q-Exponential Random Graphs: higher-order networks from simple constraints

David Dobáš,[1,*] Diego Garlaschelli,[2,3,†] and Petr Jizba[1,‡]

[1] *FNSPE, Czech Technical University in Prague, Břehová 7, 115 19, Prague, Czech Republic*
[2] *Lorentz Institute for Theoretical Physics, University of Leiden Niels Bohrweg 2, 2333 CA Leiden, The Netherlands*
[3] *IMT School for Advanced Studies Piazza S. Francesco 19, 55100 Lucca, Italy*

(Dated: May 29, 2026)

Exponential Random Graphs (ERGs) are among the most widely used network models, derived as principled least-bias graph ensembles that maximize Shannon entropy under constraints on the expected values of given structural properties. However, it has been recently (re)discovered that, in absence of additional information privileging Shannon entropy, the most agnostic inferential construction should maximize the broader class of Uffink entropies. The resulting entropy-maximizing distribution changes from the exponential (Boltzmann–Gibbs) to the so-called q-exponential one. Since maximizing Shannon entropy may produce an unjustified independence between degrees of freedom, here we investigate how the most popular ERGs with independent edges (namely, the Erdős–Rényi and configuration models) generalize to higher-order q-Exponential Random Graphs with dependent edges in the non-Shannon case, while keeping their defining constraints (number of links and degree sequence respectively) unchanged. We find features, such as a phase transition between sparse and dense regimes, that are absent in the original ERGs but typical of higher-order networks, plus novel phenomena such as richer assortativity and clustering profiles, which allow for the coexistence of link sparsity and triadic closure. These results show that higher-order networks do not necessarily require higher-order constraints, as they naturally arise from simpler ones in a framework that is even more agnostic than Shannon's.

I. INTRODUCTION

Exponential Random Graphs (ERGs) were initially proposed for social network analysis [1–5] and have been later re-derived [6] as principled unbiased models of graphs that can be successfully used, among other applications, for network reconstruction, pattern detection and combinatorial enumeration [7–11]. ERGs are defined as canonical ensembles of graphs that maximize Shannon entropy [12] under constraints on the expected value of (in principle) any desired network property. Being a family of models based on the choice of the defining constraints, ERGs are very flexible and in fact ubiquitous in network science: prominent examples include the Erdős–Rényi random graph model [13, 14] which can be obtained from a single constraint on the expected density of links in the graph, the Configuration Model [15] which corresponds to multiple constraints on the expected degrees of all nodes separately (i.e., on the expected degree sequence), and the (Degree-Corrected [16–18]) Stochastic Block Model [3, 19] which separately constrains (the degree sequence in addition to) the densities of links within and across different blocks of nodes, thus producing graphs with community structure.

ERGs permeate network science because, depending on the chosen constraints and how they are enforced, they can be used as unbiased null models to test for higher-order properties (such as reciprocity [1, 20–22], triadic closure [23, 24], network motifs [25], core-periphery structure [26], communities [27], or other statistically validated patterns [28, 29]), as structural models to replicate given features and identify structure via optimal parameter fitting [1–3, 10, 18, 30, 31], and finally as instruments for network reconstruction [32, 33], link imputation [34], and other inferential tasks [9].

A. The maximum-entropy foundation of ERGs

The reason why ERGs are so widely accepted in network science comes from their rigorous derivation from the Maximum Entropy Principle (MEP), originally introduced heuristically by Jaynes [35, 36] and later axiomatically formalized by Shore and Johnson [37, 38]. The MEP states that, given partial (aggregate) empirically available information about a (large) system — phrased in terms of expectation values (often low-order correlations or cumulants) — the least biased inference about the underlying microscopic state is a necessarily probabilistic. It is obtained by selecting the probability distribution, over all admissible microscopic states, that maximizes Shannon entropy, subject to the imposed constraints encoding the available information. In Jaynes' original formulation of the MEP, other types of constraints, such as escort means, quasi-linear means or non-inductive prior information (e.g., constraints involving the Lipschitz–Hölder regularity of probability distributions or the Hausdorff dimension of the underlying state space) are not considered. As well known [6, 11, 35] (and briefly summarized below),

*Electronic address: dobasdav@fjfi.cvut.cz
†Electronic address: garlaschelli@lorentz.leidenuniv.nl
‡Electronic address: p.jizba@fjfi.cvut.cz

the specific form of Shannon entropy implies that the resulting entropy-maximizing probability has an exponential dependence on the chosen constraints, giving (in the context of network models) ERGs their name.

But why is the MEP a rigorous construction in the first place? In turn, this comes from certain guarantees enforced via the axiomatic definition of Shannon entropy [12, 39] as the amount of uncertainty encoded in a probability distribution. In the language of random graphs relevant to this paper, if $\mathcal{G}_n$ denotes a set of graphs of interest (e.g., undirected/directed, binary/weighted, etc.) on n nodes and $P \equiv \{P(G)\}_{G\in\mathcal{G}_n}$ denotes a discrete probability distribution attaching a probability $P(G)$ to each graph $G \in \mathcal{G}_n$, then Shannon entropy is the only scalar function $S[P]$ of the elements of P that respects the following four Shannon–Khinchin (SK) axioms [39]:

- *SK1 (continuity):* $S[P]$ is continuous in the entries of P.
- *SK2 (maximality):* $S[P]$ is maximal when P is the uniform distribution over $\mathcal{G}_n$.
- *SK3 (expansibility):* $S[P]$ does not change if an extra microscopic configuration G' with zero probability $P(G') = 0$ is added to the support $\mathcal{G}_n$.
- *SK4 (additive separability):* if $P(G)$ is interpreted as a joint distribution of two random variables (here obtained by partitioning the random graph G into two non-overlapping subsets G_1 and G_2 of pairs of nodes, with $G = G_1 \cup G_2$ and $G_1 \cap G_2 = \emptyset$) with marginal distributions $Q(G_1)$ and $R(G_2)$ respectively, then $S[P]$ separates additively as

$$S[P] \;=\; S[Q] \;+\; S[R|Q]\,, \tag{1}$$

where $S[R|Q]$ is the conditional entropy defined as the mean, over $Q(G_1)$, of the entropy of the conditional distribution of G_2, given G_1. Note that in particular, if G_1 and G_2 are independent, i.e. $P(G) = Q(G_1)R(G_2)$, then $S[P]$ is simply the sum of $S[Q]$ and $S[R]$, i.e.

$$P(G) \;=\; Q(G_1)R(G_2) \;\Rightarrow\; S[P] \;=\; S[Q] \;+\; S[R]\,. \tag{2}$$

It is possible to show [39] that, up to an inessential positive prefactor, the only quantity realizing the above four SK axioms is Shannon entropy

$$S[P] \;=\; -\sum_{G\in\mathcal{G}_n} P(G)\ln P(G)\,, \tag{3}$$

which represents a cornerstone in information theory [40] and statistical physics, where it coincides with Gibbs entropic functional [41] and (upto a sign) with Boltzmann's H-function [42].

With the above properties of Shannon entropy in mind, the rationale behind the MEP is that, in presence of only aggregate information about the state of a system, the least biased probability distribution characterizing the unobserved microscopic states should be the one that encodes maximal uncertainty, besides reproducing (in expectation) the observable aggregate properties over the distribution itself [35]. This is why ERGs give a precise meaning to otherwise vague definitions such as 'random graphs with given link density', 'random graphs with given node degrees', or 'random graphs with given community structure': all such models can be rigorously understood as the least biased, maximum-entropy graph ensembles with given expected properties [6, 7, 9]. The Erdős–Rényi model [13, 14], the (canonical) Configuration Model [15] and the Stochastic Block Model [16–18] represent precisely such ensembles, respectively.

B. A generalized maximum entropy principle

In this paper, we introduce an additional element to the definition of ERGs, which descends from the recent fascinating (re)discovery that the MEP admits a more agnostic construction, based in turn on a more general family of entropy functionals with desirable inferential properties [43–45]. Indeed, while Shannon entropy is the only entropy respecting the four SK axioms as stated above, there is an important generalization that arises from relaxing the additivity requirement expressed in Eq. (1) of *SK4*. It should be noted at this point that such generalization of *SK4* has been a matter of discussion in the statistical physics literature for decades (see, e.g. [46] and references therein). This is mainly because it has been often regarded as the relaxation of Eq. (2) rather than of Eq. (1) — hence as a transition to an unavoidably non-extensive (or rather non-additive) thermodynamics [47], such as the one based on Tsallis entropy [48]

$$T_q[P] \;\equiv\; \frac{1}{1-q}\left(\sum_{G\in\mathcal{G}_n} P^q(G) - 1\right), \tag{4}$$

which indeed violates Eq. (2). However, while Eq. (3) implies Eq. (1) which in turn implies Eq. (2), there exist generalizations of Eq. (3) that violate Eq. (1), while still realizing Eq. (2), hence remaining completely additive over independent systems. The most notable such extension is Rényi entropy [49, 50], defined as

$$R_q[P] \;\equiv\; \frac{1}{1-q}\ln\sum_{G\in\mathcal{G}_n} P^q(G)\,. \tag{5}$$

To explore possible generalizations of Eq. (3) systematically, it was shown in Refs. [44, 45] that if the sum in Eq. (1) of *SK4* is replaced with the following so-called Kolmogorov–Nagumo operation[1]

$$S[P] \;=\; S[Q] \otimes_f S[R|Q]\,, \tag{6}$$

[1] Given a set M and a bijection $f^{-1} : M \mapsto N \subset \mathbb{R}$, the generalized

(arising in the context of generalized arithmetics [51, 52]), then the most general entropy consistent with the four *SK* axioms (with modified *SK4*) becomes

$$S_q^{(f)}[P] \;=\; f(U_q[P])\,, \tag{7}$$

$$U_q[P] \;=\; \left(\sum_{G\in\mathcal{G}_n} P^q(G)\right)^{\frac{1}{1-q}}\,, \tag{8}$$

where f is an arbitrary strictly increasing and positively supported function, $U_q[P]$ is what is called (for reasons that will be clear in a moment) the Uffink functional [43], and $q > 0$ is what we may call an 'entropic parameter' [53]. Indeed, the choice $f(x) = \ln x$ retrieves Rényi entropy in Eq. (5):

$$S_q^{(\ln)}[P] \;=\; R_q[P]\,. \tag{9}$$

Similarly, the choice $f(x) = \ln_q(x)$, where

$$\ln_q(x) \;\equiv\; \frac{x^{1-q}-1}{1-q}\,, \tag{10}$$

is the so-called "q-logarithm" (not to be confused with the ordinary logarithm of x to base q), retrieves Tsallis entropy in Eq. (4):

$$S_q^{(\ln_q)}[P] \;=\; T_q[P]\,. \tag{11}$$

For both entropies, the limit $q \to 1$ recovers the Shannon entropy given in Eq. (3), i.e.

$$\lim_{q\to 1} S_q^{(\ln)}[P] \;=\; \lim_{q\to 1} S_q^{(\ln_q)}[P] \;=\; S[P]\,. \tag{12}$$

The family of entropies given by Eq. (8) is quite general (see [44, 45, 53, 54]). One might nonetheless wonder whether the modified form of axiom *SK4* in Eq. (6), and hence the resulting modified entropy in Eq. (8), is merely an arbitrary construction or instead arises from a deeper mathematical consistency requirement. Crucially, it turns out that this modification is in fact necessary in order to recover the most general form of entropy consistent with another set of axioms, introduced by Shore and Johnson (*SJ*) [37, 38], which ensure the logical consistency of the MEP within Statistical Inference Theory (SIT). At this point, it is important to emphasize that, in SIT, entropy functionals serve primarily as technical tools for the unbiased assignment of probability distributions consistent with a given set of constraints. Indeed, one may argue that the maximum-entropy distribution is the primary object in SIT, while the entropy itself plays a secondary role, without an independent operational interpretation within the framework. This stands in contrast to information theory, where entropy is a primary quantity, often endowed with a clear operational meaning, for example in terms of coding theorems. Even if entropy play a secondary role in SIT, it must still satisfy a minimal set consistency conditions. Specifically, Shore and Johnson formulated their axioms not as requirements on the mathematical form of the entropy itself, but rather as conditions on the properties of the probability distribution obtained by maximizing the entropy under given constraints [37, 38]:

- *SJ1 (uniqueness):* The result (i.e., the entropy-maximizing distribution) should be unique.
- *SJ2 (invariance):* It shouldn't matter in which coordinate system one accounts for new information.
- *SJ3 (Subset independence):* It should not matter whether one treats disjoint subsets of system states in terms of separate conditional distributions or in terms of the full distribution.
- *SJ4 (System independence):* It should not matter whether one accounts for independent information (i.e., constraints) about independent systems separately in terms of marginal distributions or together in terms of a single joint distribution.
- *SJ5 (Maximality):* In absence of any prior information, the uniform distribution should be the solution.[2]

Curiously, Shore and Johnson incorrectly claimed that Shannon entropy in Eq. (3) is the only choice compatible with their axioms [37], a statement that has for a long time suggested the equivalence of the *SJ* axioms and the original *SK* ones. However, Uffink [43] later identified an error behind Shore and Johnson's initial conclusion, which was due to an unjustified hidden assumption: when considering the availability of distinct pieces of information about two subsystems of the same system, Shore and Johnson incorrectly implied that the resulting maximum-entropy joint probability factorises, *de facto* applying *SJ4* even when the independence of the two systems is not guaranteed [43, 44]. However, having disjoint pieces of information about two subsystems obviously does not guarantee that the subsystems are

Kolmogorov-Nagumo arithmetics is defined as follows:

$$\begin{aligned} x \oplus_f y &= f(f^{-1}(x) + f^{-1}(y)),\\ x \ominus_f y &= f(f^{-1}(x) - f^{-1}(y)),\\ x \otimes_f y &= f(f^{-1}(x) f^{-1}(y)),\\ x \oslash_f y &= f(f^{-1}(x) / f^{-1}(y)). \end{aligned}$$

[2] Actually, Shore and Johnson defined the maximality axiom only implicitly. Indeed, starting from the principle of minimum cross-entropy, they introduced the MEP as its equivalent in the case where the prior distribution is uniform. For this reason, even if not explicitly axiomatized, they considered the posterior P to be equal to the uniform distribution (i.e. the same as the prior) when no information is available.

independent. Unfortunately, as we discuss in detail in this paper, this is precisely what happens with the exponential probability distributions maximizing Shannon entropy, and therefore with the whole class of ERGs used so far in network science.

Remarkably, Uffink [43] showed that, if the hidden unjustified assumption of independence is dropped, then the entropy resulting from the *SJ* axioms is not necessarily the Shannon entropy, but rather the broader class of entropies given in Eq. (8) — justifying the choice of the name "Uffink functional" for the quantity $U_q[P]$. Consequently, the *SJ* axioms are, in fact, equivalent to the *modified SK* axioms [as defined via Eq. (6)] [44, 45], rather than the original ones. Both sets of axioms therefore characterize the full class of Uffink-type entropies as inferentially admissible, rather than singling out the Shannon entropy as the unique solution. Shannon entropy is the only unique solution when, in addition to the original *SJ* axioms, an additional assumption of independence is made. This also means that, in the equivalent language of *SK* axioms, the additional independence assumption corresponds to the restriction from the general Kolmogorov-Nagumo separability in Eq. (6) to the additive separability in Eq. (1), i.e. to a specific choice of f with $q \to 1$ in the Uffink functional $S_q^{(f)}[P]$ defined in Eq. (7). For this reason, q can be interpreted as a parameter controlling the degree of dependency between otherwise independent degrees of freedom in the Shannon ($q = 1$) case: values $q \neq 1$ can induce positive or negative correlations among otherwise uncorrelated variables. Similarly, as we will show in more detail, the Kolmogorov–Nagumo function f determines the scaling behavior of the entropy for uniform distributions [45].

C. From exponential to q-exponential ensembles

Accepting the extended family of entropies in Eq. (8) as the most general inferential solution implies acknowledging that the MEP should be correspondingly generalized to the search for the most general probability distribution that, given a set of input information, maximizes $S_q^{(f)}[P]$, rather than being restricted to the Shannon entropy $S[P]$ [43–45, 53, 55]. In other words, in the absence of additional information that would privilege the Shannon form, a maximally agnostic inference procedure should instead consider the broader class of Uffink-type entropies. This is because maximizing the Shannon entropy may implicitly enforce independence between degrees of freedom, which need not be justified in general. Clearly, the resulting maximum-entropy solutions therefore constitute a family of probability distributions that includes, but is not limited to, the standard exponential distributions obtained from Shannon entropy.

Note that, conveniently, the resulting generalized probability distribution is independent of f — since f is assumed to be monotonically increasing, the maximizer of $S_q^{(f)}[P]$ is unaffected by the specific choice of f — but it does depend on q. We are therefore dealing with a relatively simple, one-parameter extension of the standard exponential distribution. The ensuing generalization is known as q-exponential distribution (which indeed maximizes Tsallis entropy $T_q[P]$, Rényi entropy $R_q[R]$ and Uffink entropy $U_q[R]$ altogether [43–45, 47, 48, 53, 55]). The q-exponential function is defined as the inverse of the q-logarithm $\ln_q(x)$ defined in Eq. (10), i.e.

$$\exp_q(x) \;=\; [1 \;+\; (1-q)\,x]_+^{1/(1-q)}, \quad q \in \mathbb{R}\,, \tag{13}$$

(with $[z]_+ \equiv \max\{z, 0\}$ guaranteeing that the probability distribution is single-valued and real). In the limit $q \to 1$, the ordinary exponential function is retrieved:

$$\lim_{q\to 1} \exp_q(x) \;=\; \exp(x)\,, \tag{14}$$

consistently with the fact that, in the same limit, Shannon entropy is retrieved as previously noticed in Eq. (12).

Based on the above considerations, in this paper we make the natural logical step of introducing the q-exponential generalization of ERGs, which we denote as q-Exponential Random Graphs (q-ERGs), as the most general and agnostic class of maximum-entropy network models with given constraints. Besides the general theoretical and applied relevance of exploring q-ERGs, an aspect of particular interest is how the independence of links in Shannonian ERGs gets replaced by nontrivial dependencies in q-ERGs. This is particularly relevant in light of the recent emphasis on higher-order networks and the several attempts of introducing parsimonious models capturing them in terms, e.g. of (random) simplicial complexes and hypergraphs [56–60].

The layout of the present paper is as follows. In Sec. II, we briefly revisit some fundamentals of the MEP applied to graphs, leading to Shannonian ERG models. In particular we review how the popular Erdős–Rényi models and the Configuration Model emerge in this context as the most important models with independent edges. In Sec. III we introduce our generalized family of q-ERGs which is obtained when the MEP is extended to the Uffink entropy. In particular, we explicitly derive the generalized q-exponential graph probability distribution and introduce various technical tools and concepts that become necessary in the non-Shannon ($q \neq 1$) case, including the concept of effective Hamiltonian and a reparametrization technique that eliminates self-referentiality. In Sec. IV, we apply the q-ERG framework to obtain the natural generalization of the Erdős–Rényi model, analyzing its large-N behavior, the emergence of edge dependencies that are typical of models with higher-order constraints, and the accompanying phase transition. In Sec. V we carry out the corresponding q-exponential generalization of the Configuration Model, identifying novel results such as the possibility of enhancing the local clustering coefficients and altering the assortativity properties by varying q, while

keeping the degree sequence constrained. We finally offer our conclusions in Sec. VI.

II. TRADITIONAL SHANNONIAN ERGS

Here we briefly review the key elements of the traditional Shannon-based MEP approach [35, 37] to network ensembles, which leads to the standard ERG models [1–11] that will be relevant as our benchmarks in subsequent sections. Our particular focus will be on how the Erdős–Rényi model [13, 14] and the canonical Configuration Model [15] naturally emerge in this context. More detailed expositions can be found, e.g. in Refs. [6, 9].

The MEP approach [35] focuses on Shannon's entropy $S[P]$ [12] which, in the context of networks, is defined in Eq. (3) over the set $\mathcal{G}_n$ of all graphs with n nodes. Here $P(G)$ represents the probability of graph G in the ensemble. In the inductive inference framework, one encodes the available information in terms of K canonical constraints $\{c_i(G)\}_{i=1}^K$, each of which is a function of the considered graph G. For each constraint c_i, the available information is expressed in terms of a target value c_i^* (generally, an observed value) and the constraining procedure consists of equating the expected value (ensemble average)

$$\langle c_i \rangle \equiv \sum_{G\in\mathcal{G}_n} P(G)\, c_i(G), \quad i = 1, K\,, \tag{15}$$

with the corresponding target value c_i^*. Note that we are using angular brackets to denote averages over $P(G)$. As an additional $(K+1)$-th constraint, one has to impose the normalization condition

$$\sum_{G\in\mathcal{G}_n} P(G) = 1\,. \tag{16}$$

The resulting graph probability distribution $P(G)$ is obtained by maximizing the entropy functional, $S[P]$, subject to the aforementioned $K+1$ constraints. Introducing the associated $K+1$ Lagrange multipliers $\boldsymbol{\theta} = \{\theta_i\}_{i=1}^K$ and α, the constrained maximization of $S[P]$ is achieved by imposing the vanishing of the partial derivatives of the Lagrange functional

$$\begin{aligned}\mathcal{L}[P] \;=\; & -\sum_{G\in\mathcal{G}_n} P(G)\ln P(G) \\ & -\sum_{i=1}^{K}\theta_i\left(\sum_{G\in\mathcal{G}_n} c_i(G)P(G) - c_i^*\right) \\ & -\alpha\left(\sum_{G\in\mathcal{G}_n} P(G) - 1\right),\end{aligned} \tag{17}$$

with respect to both the Lagrange multipliers and the values of $P(G)$. The vanishing of the former derivatives generates the equations

$$c_i^* \;\equiv\; \langle c_i \rangle, \quad i \;=\; 1, K\,, \tag{18}$$

that enforce the desired constraints. The vanishing of the latter derivatives leads to the functional form of $P(G)$ which, after eliminating the multiplier α via the normalization condition [6, 9], reads

$$P(G) \;=\; \frac{1}{\mathcal{Z}(\boldsymbol{\theta})}\exp\left(-H(G,\boldsymbol{\theta})\right)\,, \tag{19}$$

where

$$H(G,\boldsymbol{\theta}) \;=\; \sum_{i=1}^{K}\theta_i c_i(G)\,, \tag{20}$$

is the so-called graph Hamiltonian. and

$$\mathcal{Z}(\boldsymbol{\theta}) \;=\; \sum_{G\in\mathcal{G}_n} e^{-H(G,\boldsymbol{\theta})}\,, \tag{21}$$

is the ensuing partition function. Therefore, in direct analogy with statistical physics, the form of $P(G)$ mimics the Boltzmann distribution for the canonical ensemble.

Once the probability $P(G)$ is determined, it can be used to express the expected value $\langle c_i \rangle$ of each constraint $c_i(G)$ as a function of the parameters $\boldsymbol{\theta}$. Alternatively, the same result is obtained via the derivative

$$\langle c_i \rangle \;=\; -\frac{\partial}{\partial\theta_i}\ln\mathcal{Z}(\boldsymbol{\theta})\,. \tag{22}$$

Finally, the values of the parameters $\boldsymbol{\theta}$ are obtained by equating the expected value of each constraint with the corresponding target value into Eq. (18), which in general represents a system of K nonlinear coupled equations in the K unknowns $\{\theta_i\}_{i=1}^K$. Various techniques to identify the corresponding solution have been introduced [10, 30, 61, 62].

The central goal in the analysis of ERGs is the explicit calculation of the partition function $\mathcal{Z}(\boldsymbol{\theta})$ for a given choice of the constraints $\{c_i(G)\}_{i=1}^K$. Depending on the definition of the latter, this task can be more or less easy (or even impossible) [9, 10]. For our purposes here, we will consider simple additive choices of the constraints (which are also the most widespread choices adopted in the analysis of real-world networks) that lead to explicit forms of $\mathcal{Z}(\boldsymbol{\theta})$. Importantly, these choices highlight the resulting independence between links in the graph, that follows from the Shannonian ($q = 1$) assumption. Therefore these models represent the cleanest benchmarks against which the novelty of q-ERGs will emerge in our subsequent analysis in Sec. III. Concretely, in the rest of the paper we consider examples of ensembles obtained by choosing constraints on simple undirected graphs (without self-loops) on n nodes. From now on, $\mathcal{G}_n$ will therefore denote the set of all such graphs. Note that the cardinality of this set is $|\mathcal{G}_n| = 2^N$ where $N = \binom{n}{2} = n(n-1)/2$ is the number of distinct pairs of n nodes (which coincides with maximum number of links that can be realized).

A. Erdős–Rényi model

We first consider the case where there is only one ($K = 1$) constraint, namely the mean value of the total number of links in the graph, which we denote as $c_1(G) = L(G)$. It is convenient to introduce the adjacency matrix of a binary undirected graph G, which is a symmetric matrix A with elements $a_{ij} = 1$ if there is a link between nodes i and j, and $a_{ij} = 0$ otherwise (with $a_{ii} = 0$). For simple undirected graphs, the number of links in G is given by

$$L(G) \;=\; \sum_{i=1}^{n}\sum_{i<j} a_{ij}\,, \tag{23}$$

and we set the expected constraint to the target value $c_1^* = L^*$ as in the general prescription given by Eq. (18):

$$L^* \;\equiv\; \langle L\rangle \;=\; \sum_{G\in\mathcal{G}_n} P(G)L(G)\,. \tag{24}$$

This leads to the graph Hamiltonian

$$H(G,\theta) \;=\; \theta L(G) \;=\; \theta\sum_{i=1}^{n}\sum_{j\neq i} a_{ij}\,, \tag{25}$$

and partition function

$$\mathcal{Z}(\theta) \;=\; \sum_{G\in\mathcal{G}_n} e^{-\theta L(G)}. \tag{26}$$

Changing notation from a sum over G to a sum over all possible assignments $\{a_{ij}\}$ of the corresponding adjacency matrix entries, we can evaluate the partition function explicitly [6, 9] as

$$\begin{aligned}\mathcal{Z}(\theta) \;&=\; \sum_{\{a_{ij}\}} e^{-\theta\sum_{i=1}^{n}\sum_{j<i}a_{ij}} \;=\; \sum_{\{a_{ij}\}}\prod_{i=1}^{n}\prod_{j<i} e^{-\theta a_{ij}}\\ &=\; \prod_{i=1}^{n}\prod_{j<i}\sum_{a_{ij}=0,1} e^{-\theta a_{ij}} \;=\; \prod_{i=1}^{n}\prod_{j<i}(1+e^{-\theta})\\ &=\; (1+e^{-\theta})^N\,. \end{aligned}\tag{27}$$

The ensuing probability of the entire graph G then factorizes over pairs of nodes as follows:

$$\begin{aligned}P(G) \;&=\; \frac{e^{-\theta\sum_{i=1}^{n}\sum_{j<i}a_{ij}}}{\mathcal{Z}(\theta)}\\ &=\; \prod_{i=1}^{n}\prod_{j<i}\frac{e^{-\theta a_{ij}}}{1+e^{-\theta}} \;=\; \prod_{i=1}^{n}\prod_{j<i} P(a_{ij})\,, \end{aligned}\tag{28}$$

where $P(a_{ij}) = e^{-\theta a_{ij}}/(1+e^{-\theta})$ is the probability of the link between nodes i and j being either present ($a_{ij} = 1$) or absent ($a_{ij} = 0$) [6, 9].

Thus, this model is a model of a random graph with independent links — each link having the same probability

$$\langle a_{ij}\rangle \;=\; P(1) \;=\; \frac{e^{-\theta}}{1+e^{-\theta}} \;=\; \frac{1}{1+e^{\theta}} \;\equiv\; p\,, \tag{29}$$

of being present, which is equivalent to modeling $\binom{n}{2}$ i.i.d. Boolean variables. Such model of graphs with independent links, with each link having the same probability p, is precisely the celebrated *Erdős–Rényi* (ER) model [13, 14]. The connection probability p is nothing but a reparametrization of the Lagrange multiplier θ that controls the expected total number of links

$$\langle L\rangle \;=\; \sum_{i=1}^{n}\sum_{i<j}\langle a_{ij}\rangle \;=\; pN\,, \tag{30}$$

or, equivalently, the expected overall link density $\langle L\rangle/N = p$, see [6, 9]. Indeed, the value of p enforcing the desired target number of links

$$L^* \;=\; pN\,, \tag{31}$$

as required in Eq. (24), is $p = L^*/N$, i.e. it equals the target link density.

The complete homogeneity of the ER model fails in replicating the complex features of real-world networks, such as the broad distribution of the number of links (degree) per node and the asymptotic coexistence of *sparsity* (vanishing link density) and *clustering* (finite average fraction of realized triangles around nodes) for a large number n of nodes. For this reason, while the ER model remains a useful benchmark of homogeneous graphs, more complicated ERGs with node-specific constraints are used in network science, as we now consider in our next example.

B. Canonical Configuration Model

In the canonical Configuration Model (CM), the expected degree (number of links) of each node is constrained [6, 9]. The number of constraints is therefore $K = n$. In terms of the adjacency matrix, the degree $k_i(G)$ of node i in graph G is defined as

$$k_i(G) \;=\; \sum_{j\neq i} a_{ij}, \quad i = 1, n\,. \tag{32}$$

This leads to a graph Hamiltonian with n constraints $c_i = k_i$ ($i = 1, n$):

$$\begin{aligned}H(G,\boldsymbol{\theta}) \;&=\; \sum_{i=1}^{n}\theta_i k_i(G) \;=\; \sum_{i=1}^{n}\sum_{j\neq i}\theta_i a_{ij}\\ &=\; \sum_{i=1}^{n}\sum_{j<i}(\theta_i+\theta_j)a_{ij}\,. \end{aligned}\tag{33}$$

The corresponding partition function is a straightforward generalization of Eq. (27) and is given by [6, 9]

$$\mathcal{Z}(\boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j<i}\left(1+e^{-\theta_i-\theta_j}\right) . \tag{34}$$

The graph probability still factorizes as

$$P(G) = \prod_{i=1}^{n}\prod_{j<i}\frac{e^{-(\theta_i+\theta_j)a_{ij}}}{1+e^{-(\theta_i+\theta_j)}} = \prod_{i=1}^{n}\prod_{j<i}P_{ij}(a_{ij}) , \tag{35}$$

where $P_{ij}(a_{ij}) = e^{-(\theta_i+\theta_j)a_{ij}}/(1+e^{-(\theta_i+\theta_j)})$ is the probability of the link between nodes i and j being either present ($a_{ij}=1$) or absent ($a_{ij}=0$). However, the link probabilities are now heterogeneous:

$$\begin{aligned}\langle a_{ij}\rangle &= P_{ij}(1) = \frac{e^{-\theta_i-\theta_j}}{1+e^{-\theta_i-\theta_j}} \\ &= \frac{x_i x_j}{1+x_i x_j} \equiv p_{ij} ,\end{aligned} \tag{36}$$

where we have employed the conventional reparametrization from $\boldsymbol{\theta}$ to the so-called fitness parameters $\boldsymbol{x}$ via the prescription $x_i = e^{-\theta_i}$ [6, 9].

Now, equating the expected degrees to the target ones following the general requirement of Eq. (18) leads to the following n coupled nonlinear equations for the n unknown parameters [9, 31]:

$$k_i^* \equiv \langle k_i\rangle = \sum_{j\neq i} p_{ij} = \sum_{j\neq i}\frac{x_i x_j}{1+x_i x_j}, \quad i = 1,n , \tag{37}$$

where k_i^* is the desired target value of the degree of node i. Notably, these equations coincide with a maximum-likelihood condition on the parameters $\boldsymbol{x}$ [9]. The value of $\boldsymbol{x}$ providing a solution to the above system of equations is unique, and efficient algorithms to find it for large networks where k_i^* is prescribed for each i separately (typically, when it represents the empirical degree observed in a real-world network [9, 31]) exist [62]. With these parameter values, the CM can be used as a null model of a given network with degrees $\{k_i^*\}_{i=1}^n$, e.g. to look for possible higher-order patterns in terms of statistically significant deviations from the maximum-entropy ensemble [9, 31].

There is however also a different, 'reverse' approach, where the parameters $\boldsymbol{x}$ are treated as free parameters (e.g. the values $\{x_i\}_{i=1}^n$ can be sampled i.i.d. from a probability distribution of choice) that can be varied in order to explore the parameter space of the model and study the resulting value of the expected degrees [6] and of other topological properties of the network, without a specific target network in mind. Still, one might want to restrict the search in order to compare outcomes obtained by different values of $\boldsymbol{x}$ while keeping certain overall properties, such as the overall link density, fixed[3]. A particularly convenient strategy (that we will adopt later) is to rewrite $x_i \equiv \sqrt{z}s_i$ where s_i is still a node-specific parameter, while z is a global one. In this way it is possible to sample $\{s_i\}_{i=1}^n$ i.i.d. from any desired distribution, while adjusting z to keep the average link density in the network fixed to a prescribed value, independent of that distribution. This means that the connection probability can be rewritten as

$$p_{ij} = \frac{z s_i s_j}{1+z s_i s_j} , \tag{38}$$

where z is determined, conditionally on the realized values of $\{s_i\}_{i=1}^n$, by enforcing a constant expected link density $c^* = \langle L\rangle/N$ as follows:

$$c^* \equiv \langle c\rangle = \frac{1}{N}\sum_{i=1}^{n}\sum_{j<i}\frac{z s_i s_j}{1+z s_i s_j} . \tag{39}$$

For fixed $\{s_i\}_{i=1}^n$, the above equation coincides with a maximum-likelihood condition for z [63]. With the above strategy, it is possible to vary $\{s_i\}_{i=1}^n$ and re-fit z to effectively explore different fitness assignments $\{x_i\}_{i=1}^n$ that lead to exactly the same expected link density. This is also the standard approach in techniques of network reconstruction from partial infromation [32, 33] that take $\{s_i\}_{i=1}^n$ as input in terms of some observable node feature, while requiring only the empirical knowledge of the overall link density c^* in order to construct maximum-entropy ensembles for network inference based on link probabilities in the form of Eq. (38). We will use the above approach to keep the network density fixed while exploring the q-exponential generalization of the CM in the next section.

The CM model improves upon the ER model by allowing the degree distribution of any real network to be replicated in expectation. However, it still cannot replicate the coexistence of sparsity (i.e. the average node degree remaining finite as the number n of vertices becomes larger and larger) and clustering (i.e. a finite value of the local clustering coefficient). For this reason, models with independent links (like the ER and CM) are believed to be unable to replicate higher-order properties besides those included in the constraints. This has led to the expectation that higher-order constraints need be included in the construction of the maximum-entropy probability in order to replicate those additional properties. However, in what follows we show that this conclusion is not correct when considering more general entropies. The same simple constraints as those defining the ER and

[3] Note that, in order to keep the expected overall link density fixed for different choices of the fitness vector $\boldsymbol{x}$, it is not appropriate to demand that the mean of the distribution from which the entries of $\boldsymbol{x}$ are sampled i.i.d. (hence the expected average of the n fitness values) is kept fixed. This is due to the nonlinearity of the relation between fitness x_i and expected degree $\langle k_i\rangle$ in Eq. (37). For this reason, the approach described here and based on the global parameter z is necessary.

CM can indeed lead to link dependencies, which in turn can generate higher-order properties, including a finite clustering in the sparse regime.

III. q-EXPONENTIAL RANDOM GRAPHS

We now come to our main contribution, by moving beyond the Shannon case considered in Sec. II and considering the extended family of Uffink entropies given by Eq. (8) as the object of the generalized MEP. As already anticipated, the main result of the maximization of the Uffink functional will be the replacement of the inverse exponential distribution in Eq. (19) with the inverse q-exponential function defined in Eq. (13). This operation will define the new family of q-ERGs in terms of a certain distribution that we denote as $P_q(G)$. To show explicitly how $P_q(G)$ emerges from the maximization of any of the entropies in Eq. (8) for general q, we need to pick a specific choice of f. Concretely, we select $f(x) = \ln_q(x)$ given by Eq. (10), which leads to the Tsallis entropy functional $T_q[P]$ in Eq. (4) [48], even though the result will in the end not depend on f (and on the resulting from of entropy) because, as we mentioned, the maximizer of a function is also the maximizer of any monotonic function f of that function.

As in the Shannonian case, a key step is selecting the K constraints $\{c_i(G)\}_{i=1}^K$ and imposing definite values for their expected values. However, while in the $q = 1$ case Eq. (15) provides a unique natural definition of the expected value $\langle c_i\rangle$ of in terms of the distribution $P(G)$ itself, multiple definitions of q-expected values $\langle c_i\rangle_q$ exist in the $q \neq 1$ case, all reducing to the standard $\langle c_i\rangle$ in the limit $q \to 1$. In particular, three averaging procedures have been used more often in the context of q-exponentials. The so-called q-average, also known as the Curado–Tsallis weighted mean (or 2nd version of thermostatistics) [64], replaces $P_q(G)$ with $P_q^q(G)$ in the definition of the expected constraints and offers certain formal advantages. For example, it allows one to set up a formalism that closely emulates the formalism known from conventional thermodynamics [65]. However, the q-average has a controversial status because it violates normalization ($\sum_G P_q^q(G) \neq 1$ for $q \neq 1$) and does not have an easy operational meaning. Another procedure is the 3rd version of thermostatistics, which replaces $P_q(G)$ with the so-called escort distribution $P_q^q(G)/\sum_{G'} P_q^q(G')$ and allows for a natural generalization of the maximum-likelihood principle to the estimation of the entropic parameter q [53], restoring the correspondence between entropy and negative log-likelihood. Conversely, the conventional linear averaging procedure based on $P_q(G)$ (also known as the Bashkirov's 1st version of thermostatistics) is conceptually cleanest from a statistical inference point of view [44, 45]. In addition, it aligns well with network science, where standard (linear) ensemble averages are routinely used in ERGs as shown above. For this reason, we adopt standard averaging in our subsequent analysis and simply set

$$\langle c_i\rangle_q \equiv \langle c_i\rangle = \sum_{G\in\mathcal{G}_n} P_q(G)\, c_i(G), \quad i = 1, K\,, \tag{40}$$

in analogy with Eq. (15). In any case, all the three choices (plus additional ones) can be made equivalent to each other by an appropriate change of q (e.g. via the duality between q and $2-q$) and/or the way the Lagrange multipliers $\{\theta_i\}_{i=1}^K$ are estimated [53, 66].

In order to maximize the entropy $T_q[P]$ under the K constraints $\langle c_i\rangle = c_i^*$ $(i = 1, K)$ plus the additional constraint on the normalization of P as in Eq. (16), we define the q-Lagrange functional $\mathcal{L}_q[P]$ which is a direct generalization of the Shannonian $\mathcal{L}[P]$ in Eq. (17):

$$\begin{aligned}\mathcal{L}_q[P] &= \frac{1}{q-1}\left(1 - \sum_{G\in\mathcal{G}_n} P^q(G)\right) \\ &\quad - \sum_{i=1}^K \theta_i \left(\sum_{G\in\mathcal{G}_n} c_i(G)P(G) - c_i^*\right) \\ &\quad - \alpha\left(\sum_{G\in\mathcal{G}_n} P(G) - 1\right).\end{aligned} \tag{41}$$

Optimizing this functional (see Appendix A for the full derivation) yields the q-exponential graph probability

$$\begin{aligned}P_q(G) &= \frac{\left[\exp_q\left(\frac{1}{q\mathcal{Z}_q^{1-q}(\boldsymbol{\theta})}\sum_{i=1}^K \theta_i(c_i(G) - \langle c_i\rangle)\right)\right]^{-1}}{\mathcal{Z}_q(\boldsymbol{\theta})} \\ &= \frac{\exp_{2-q}\left(\frac{-1}{q\mathcal{Z}_q^{1-q}(\boldsymbol{\theta})}\sum_{i=1}^K \theta_i(c_i(G) - \langle c_i\rangle)\right)}{\mathcal{Z}_q(\boldsymbol{\theta})}\,,\end{aligned} \tag{42}$$

where the partition function equals $\mathcal{Z}_q(\boldsymbol{\theta}) = U_q[P_q]$ and, in the second line, we have have exploited the convenient duality relation $\exp_q(-x) = \exp_{2-q}^{-1}(x)$ (see Appendix A).

Although Eq. (42) is self-referential (both the partition function $\mathcal{Z}_q$ and the constraints $\langle c_i\rangle$ appear inside the exponent), this problem can be resolved in a standard way by reparametrizing the Lagrange multipliers $\theta_i \to \tilde{\theta}_i$ [67] (see Appendix A for details). Although the parameters $\tilde{\theta}_i$ are no longer the Lagrange multipliers of the original problem, they are still supposed to be fitted according to the constraints. To simplify notation, we henceforth drop the extra symbol and denote the redefined parameters simply as θ_i and the corresponding partition function as $\mathcal{Z}_q(\boldsymbol{\theta})$. The general q-exponential probability can then be shown to take the compact form

$$P_q(G) = \frac{1}{\mathcal{Z}_q(\boldsymbol{\theta})}\exp_{2-q}(-H(G,\boldsymbol{\theta}))\,, \tag{43}$$

where $H(G,\boldsymbol{\theta})$ has the same definition as in Eq. (20) for the $q = 1$ case. Equation (43) finally generalizes ERGs to a one-parameter family of q-ERGs, hence representing the formal solution to the most agnostic inferential construction of graph ensembles with soft constraints.

A. Effective Hamiltonian

Clearly, when $q \to 1$ Eq. (43) reduces to the exponential distribution in Eq. (19), and q-ERGs reduce to ordinary ERGs. In order to better understand the differences (with respect to the Shannonian case) introduced by a value $q \neq 1$ in the q-exponential framework, we introduce the *effective Hamiltonian* $H_q^{\rm eff}(G,\boldsymbol{\theta})$ defined implicitly by

$$P_q(G) \equiv \frac{1}{\mathcal{Z}_q(\boldsymbol{\theta})} \exp\big(-H_q^{\rm eff}(G,\boldsymbol{\theta})\big) . \tag{44}$$

The quantity $H_q^{\rm eff}(G,\boldsymbol{\theta})$ is defined in such a way to provide us with an indication of which 'effective constraints' one should use in the Shannonian ERG formalism in order to produce a graph probability of the same form (up to a proportionality constant) as the ordinary exponential probability derived in Eq. (19) for the $q = 1$ case. One might then see this operation as a mapping from the original set of constraints used to define a q-ERG to a different set of effective constraints necessary to define an equivalent ordinary ERG producing the same probability distribution $P_q(G)$.

Equation (44) is equivalent to setting

$$\begin{aligned} H_q^{\rm eff}(G,\boldsymbol{\theta}) &\equiv -\ln \exp_{2-q}(-H(G,\boldsymbol{\theta})) \\ &= \ln \exp_q H(G,\boldsymbol{\theta}) \\ &= \frac{1}{1-q} \ln\big[1 + (1-q)H(G,\boldsymbol{\theta})\big]_+ \\ &= \sum_{k=1}^{+\infty} \frac{(q-1)^{k-1}}{k} H^k(G,\boldsymbol{\theta}) . \end{aligned} \tag{45}$$

where in the last line we require $|1-q| \leq |H(G,\boldsymbol{\theta})|^{-1}$ since we have employed the Mercator series expansion

$$\ln(1+x) = \sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{k} x^k \tag{46}$$

of the logarithm, which is valid for $|x| \leq 1$ and $x \neq -1$. In particular, considering values of q sufficiently close to $q = 1$ (this is the most relevant regime, as we show in the next sections), we get the following expansion for the effective Hamiltonian:

$$\begin{aligned} H_q^{\rm eff}(G,\boldsymbol{\theta}) &= H(G,\boldsymbol{\theta}) \\ &\quad + \tfrac{1}{2}(1-q)H^2(G,\boldsymbol{\theta}) \\ &\quad + \tfrac{1}{3}(1-q)^2 H^3(G,\boldsymbol{\theta}) \\ &\quad + \tfrac{1}{4}(1-q)^3 H^4(G,\boldsymbol{\theta}) \\ &\quad + \mathcal{O}\big((1-q)^4\big) . \end{aligned} \tag{47}$$

We thus realize that, as seen from an ordinary ERG viewpoint, the q-ERG generates the additional terms

$$\begin{aligned} \Delta H_q(G,\boldsymbol{\theta}) &\equiv H_q^{\rm eff}(G,\boldsymbol{\theta}) - H(G,\boldsymbol{\theta}) \\ &= \sum_{k=2}^{+\infty} \frac{(q-1)^{k-1}}{k} H^k(G,\boldsymbol{\theta}) , \end{aligned} \tag{48}$$

besides those that would normally be present in the Shannon ERG case. These terms are produced by all possible k-th powers of $H(G,\boldsymbol{\theta})$ with $k > 1$. It is however important to note that there are no additional Lagrange multipliers coupled with those extra terms: the only additional parameter is the single, global entropic parameter q, whose value couples with the K native Lagrange multipliers $\boldsymbol{\theta}$ in determining $P_q(G)$ and automatically dictates the relative importance of any power $H^k(G,\boldsymbol{\theta})$ appearing in $H_q^{\rm eff}(G,\boldsymbol{\theta})$. Therefore the new effective quantities introduced by $\Delta H_q(G,\boldsymbol{\theta})$ should not in fact be interpreted as new tunable 'constraints', but rather as new structural features that will be unavoidably promoted in the network ensemble, with a strength depending on the value of q, once the K original constraints $\{\langle c_i \rangle\}_{i=1}^K$ are enforced.

A direct consequence of the presence of the extra ($k > 1$) terms in the expansion (45) for $H_q^{\rm eff}(G,\boldsymbol{\theta})$ is that, when the defining constraints in the Hamiltonian $H(G,\boldsymbol{\theta})$ of the q-ERG (i.e., the first term $k = 1$) are sums of elements of the adjacency matrix (e.g. the total number of links or the degree sequence, as in the ER and CM examples we considered in Sec. II A and II B respectively), links will no longer be independent as we have found for the standard ERG, because the q-exponential of $H(G,\boldsymbol{\theta})$ generates effective higher-order link interactions at every order in $(q-1)$. Specifically, in such models the quadratic term $\frac{1}{2}(q-1)H^2(G,\boldsymbol{\theta})$ will introduce pairwise link–link interactions, the cubic term $\frac{1}{3}(q-1)^2 H^3(G,\boldsymbol{\theta})$ will introduce three-link interactions, and so on at every subsequent order $k^{-1}(q-1)^{k-1}H^k(G,\boldsymbol{\theta})$. Each such order involves products of k link variables $a_{i_1 j_1} a_{i_2 j_2} \cdots a_{i_k j_k}$, thereby coupling links that are otherwise statistically independent in the ordinary ($q = 1$) ERG case. Hence, *even without modifying the constraints*, the mere departure from Shannon entropy ($q \neq 1$) transforms an edge-independent model into a genuinely higher-order one, coupling arbitrarily many links. In the rest of the paper, we investigate in detail these q-ERGs and indeed find novel and much more general phenomenology with respect to what ordinary ERGs feature.

IV. q-EXPONENTIAL ERDŐS–RÉNYI MODEL

We now work out in detail the simplest possible q-ERG, which is the direct analogue of the ER model considered in Sec. II A. The resulting model, which we denote as the q-exponential ER (q-ER) model, is still defined by the same single ($K = 1$) constraint on the expected to-

tal number of links $\langle L\rangle$, which is set equal to a target value L^*, and the Hamiltonian is therefore the same as in Eq. (25). Using Eq. (43), the resulting maximum q-entropy probability distribution is given by

$$P_q(G) = \frac{1}{\mathcal{Z}_q(\theta)} \exp_{2-q}\left(-\theta L(G)\right) = \frac{1}{\mathcal{Z}_q(\theta)} \exp_{2-q}\Big(-\theta \sum_{i=1}^{n}\sum_{j<i} a_{ij}\Big), \quad (49)$$

with partition function

$$\mathcal{Z}_q(\theta) = \sum_{\{a_{ij}\}} \exp_{2-q}\Big(-\theta \sum_{i=1}^{n}\sum_{j<i} a_{ij}\Big). \quad (50)$$

The significant departure from the Shannonian case lies in the fact that, even though the Hamiltonian is still a sum of the pairwise elements a_{ij}, the maximum-entropy probability $P_q(G)$ no longer factorizes into a product of probabilities of individual links as in Eq. (28), because the $\mathcal{Z}_q(\theta)$ does not factorize either: indeed, the $(2-q)$-exponential of the sum inside Eq. (50) is not the product of the $(2-q)$-exponentials, unless $q=1$. This means that links are not independent random variables anymore, and the q-ER model is actually a higher-order network model with couplings among any number of links. To confirm this, we notice that the effective Hamiltonian introduced in Eq. (44) reads

$$\begin{aligned} H_q^{\text{eff}}(G,\theta) &= H(G,\theta) + \Delta H_q(G,\theta) \quad (51)\\ &= \theta \sum_{i=1}^{n}\sum_{j<i} a_{ij} \\ &\quad + \tfrac{1}{2}(q-1)\theta^2 \left(\sum_{i=1}^{n}\sum_{j<i} a_{ij}\right)^2 \\ &\quad + \tfrac{1}{3}(q-1)^2\theta^3 \left(\sum_{i=1}^{n}\sum_{j<i} a_{ij}\right)^3 \\ &\quad + \mathcal{O}\big((q-1)^3\big). \end{aligned}$$

We therefore see that each term of order $(q-1)^k$ generates a coupling between a number $k+1$ of pairs of nodes. In the next sections we will explore the consequences of these couplings and their relation with ERGs with higher-order constraints.

Before doing that, we notice that the (modified) Lagrange multipliers appearing in $P_q(G)$ are obtained by imposing the matching condition equivalent to the one in Eq. (24) with $P(G)$ replaced by $P_q(G)$, i.e.

$$L^* \equiv \langle L\rangle = \sum_{G\in\mathcal{G}_n} P_q(G)L(G), \quad (52)$$

where now the matching value of θ will depend on q [53], and will reduce to the ordinary value discussed in Sec. II A when $q\to 1$. Note that, even though links are all dependent on one another, the expected value $\langle a_{ij}\rangle$ still represents the probability of connection between nodes i and j; however, this is now interpreted as the *marginal* connection probability, unconditional on the realization of other links, which are all dependent on a_{ij}. Since $P_q(G)$ depends on the adjacency matrix $\{a_{ij}\}$ only through the total number of links $L(G)=\sum_{i<j} a_{ij}$, it is invariant under any permutation of links, so the value of p_{ij} is independent of i,j, as in the traditional ER model:

$$p_{ij} \equiv \langle a_{ij}\rangle = \frac{\langle L\rangle}{N} = \frac{L^*}{N}, \qquad \forall i\neq j. \quad (53)$$

The above relation is useful to compare the properties of the q-ER model with those of the corresponding ordinary ER model for the same expected link density L^*/N.

A. Relation with ERGs with higher-order constraints

We now go back to the effective Hamiltonian in Eq. (51) and compare it with known ERGs with higher-order constraints.

For instance, the term of order $q-1$ can be rewritten as $(\sum_{i=1}^n\sum_{j<i} a_{ij})^2 = \sum_{i=1}^n\sum_{j<i}\sum_{k=1}^n\sum_{l<k} a_{ij}a_{kl}$, which generates 2-link couplings of the form $a_{ij}a_{kl}$. A subset of such couplings is present, for instance, in the so-called 2-star model [68, 69], a popular ERG obtained by enforcing two global constraints in the network: the total number of links $\sum_{i=1}^n\sum_{j<i} a_{ij}$ (the lowest-order constraint) and the total number $\sum_{i=1}^n\sum_{j<i}\sum_{k<i} a_{ij}a_{ik}$ (with $k\neq j$) of so-called 2-stars, which is one of the simplest higher-order constraints. Each of the two constraints is controlled by a separate Lagrange multiplier.

Similarly, the term of order $(q-1)^2$ can be rewritten as $\sum_{i=1}^n\sum_{j<i}\sum_{k=1}^n\sum_{l<k}\sum_{u=1}^n\sum_{v<u} a_{ij}a_{kl}a_{uv}$, which generates 3-link couplings of the form $a_{ij}a_{kl}a_{uv}$. A subset of these couplings is present in the edge-triangle or Strauss model [2, 10, 70], another popular ERG obtained by enforcing the total number of links $\sum_{i=1}^n\sum_{j<i} a_{ij}$ and the total number $\sum_{i=1}^n\sum_{j<i}\sum_{k<j<i} a_{ij}a_{jk}a_{ki}$ of triangles. Here as well, the lower-order and the higher-order constraints are controlled by two independent Lagrange multipliers.

The 2-star model and the edge-triangle model are known to be characterized by a phase transition [2, 10, 68–70]: for certain combinations of the values of the two Lagrange multipliers, the ERG probability $P(G)$ is concentrated over a set of configurations where the constraints are close to their target expected values; however, for other combinations $P(G)$ is concentrated over multiple sets of configurations, each set featuring values of the constraints that are either much smaller or much larger than the target values, while still replicating the correct

number in expectation. This result implies that, in individual realizations of the ERG, not all combinations of the values of the defining constraints can be realized. Certain combinations can only be replicated upon averaging over realizations that are individually very different from the desired configuration.

In the next sections we will show that a qualitatively similar phase transition is present in the q-ER model as well, however with important differences: unlike the 2-star model and the edge-triangle model, the q-ER model features couplings of all possible (infinite) orders; the introduction of these couplings is not *ad hoc* or arbitrary, because the only chosen constraint is the total number of links, all the other quantities in $\Delta H_q(G,\theta)$ emerging spontaneously as a consequence of having relaxed the entropic parameter to values $q \neq 1$; the higher-order couplings need not be tuned independently of one another, since they are all governed by specific combinations of the two parameters θ and q.

B. Thermodynamic limit

We now study the asymptotic behavior of the q-ER probability distribution in Eq. (49) and the ensuing partition function for large N. The limit $N \to \infty$ corresponds to the 'thermodynamic limit' where the size of the graph becomes infinite, while certain appropriately rescaled properties converge to limiting values.

We adopt a methodology familiar from the analysis of the Asymptotic Equipartition Property in information theory [40] and focus on the large-N behavior of the quantity $(\ln P)/N$. To this end, we start with a related quantity known as the *free entropy density*

$$\Phi_q(\theta, N) \;=\; \frac{1}{N} \ln \mathcal{Z}_{q,N}(\theta)\,, \tag{54}$$

which an analogue of a Massieu function from equilibrium thermodynamics. From now on, we will use the notation $\mathcal{Z}_{q,N}(\theta)$, since we are interested in the large N behavior and the partition function is explicitly N dependent (in eq. (50) we are summing all the N link variables a_{ij}, thus N is effectively the system size). In appendix B, we demonstrate that, in the limit of large N, the value of (54) is determined by the maximum of yet another quantity, which we will refer to as the *effective free entropy density* $\phi_q(c,\theta,N)$:

$$\lim_{N\to\infty} \Phi_q(\theta, N) \;=\; \lim_{N\to\infty} \max_{c\,\in\,[0,1]} \phi_q(c,\theta,N)\,, \tag{55}$$

where

$$\phi_q(c,\theta,N) \;=\; H_b(c) \;-\; \frac{1}{(1-q)N} \ln\left[1 + (1-q)N\theta c\right]\,, \tag{56}$$

and $H_b(c) = -c\ln c - (1-c)\ln(1-c)$ is the binary entropy.

We now stress that, when considering the thermodynamic limit $N \to \infty$, one should crucially consider how the entropy parameter q scales with N. We first consider q constant and finite as N grows, and show that this assumption has to be revised for any nontrivial result to emerge.

1. *Fixed q*

Assuming fixed θ and q as $N \to \infty$, we have

$$\lim_{N\to\infty} \frac{1}{(1-q)N} \ln\left(1 + (1-q)N\theta c\right) \;=\; 0\,. \tag{57}$$

Thus, the second term in (56) becomes negligible for large N, and the behavior of Φ_q and $\mathcal{Z}_{q,N}(\theta)$ is determined entirely by the binary entropy term. In the large N limit (for fixed θ and q), we thus obtain for the free entropy density

$$\lim_{N\to\infty} \Phi_q(\theta, N) \;=\; \max_{c\,\in\,[0,1]} H_b(c) \;=\; H_b(1/2)\,, \tag{58}$$

where we used the fact that the binary entropy $H_b(c)$ is maximized by $c_{\max} = \arg\max_c H_b(c) = \frac{1}{2}$. Consequently, we have

$$\lim_{N\to\infty} \frac{\ln \mathcal{Z}_{q,N}(\theta)}{N} \;=\; H_b(1/2)$$
$$\Leftrightarrow \quad \mathcal{Z}_{q,N}(\theta) \;\asymp\; e^{N H_b(1/2)} \;=\; 2^N\,, \tag{59}$$

where we have employed the notation for an asymptotic equivalence in the exponential sense, namely

$$a_N \asymp b_N \quad \Leftrightarrow \quad \lim_{N\to\infty} \frac{1}{N} \ln a_N \;=\; \lim_{N\to\infty} \frac{1}{N} \ln b_N\,. \tag{60}$$

Similarly, we also show in Appendix B that

$$P\left(\frac{L(G)}{N} = c\right) \;\asymp\; e^{-N[H_b(1/2)\;-\;H_b(c)]}\,. \tag{61}$$

This shows that the probability is maximal at $c = 1/2$ and decays exponentially away from this value. Such behavior exemplifies the *large deviation principle* [71], with the *rate function*

$$I(c) \;\equiv\; H_b(1/2) \;-\; H_b(c)\,, \tag{62}$$

determining the exponential decay. The unique minimum $I(1/2) = 0$ implies that only states with $c = 1/2$ retain non-negligible probability as $N \to \infty$, a phenomenon known as the *concentration of measure*.

This reveals a striking *degeneracy*: keeping q and θ fixed (the latter condition corresponds to the dense graph case), the large-N free entropy density converges to the universal value $H_b(1/2) = \ln 2$, identical to that of a completely unbiased Erdös–Rényi graph with link probability 1/2. In other words, both θ and q become irrelevant in

the thermodynamic limit at fixed values — the model loses all memory of its parameters and concentrates all probability on the maximally random configuration. This is clearly pathological from the modelling perspective, and motivates the question of whether and how this degeneracy can be overcome.

Let us now see how this degenerate case compares with the Shannonian one, where the probability of observing a graph with given density c is given by

$$P\left(\frac{L(G)}{N}=c\right) \;=\; \binom{N}{Nc}\frac{1}{\mathcal{Z}^N(\theta)}\,e^{-\theta Nc}\,. \tag{63}$$

The effective free entropy density then takes the form

$$\phi_1(c,\theta,N) \;\equiv\; \phi_1(c,\theta) \;=\; H_b(c) \;-\; \theta c\,. \tag{64}$$

Notably, it is independent of N, if the contribution of the constraint term θc remains finite (this is the dense graph limit serving as a reference for the dense $q \neq 1$ case considered above). Extremizing with respect to c yields a single stationary point at

$$c_{\max} \;=\; \frac{1}{1\;+\;e^{\theta}}\,, \tag{65}$$

in agreement with eq. (29). The probability then reads

$$P\left(\frac{L(G)}{N}=c\right) \;\asymp\; e^{-N[\phi_1(c_{max},\theta)\;-\;\phi_1(c,\theta)]}\,. \tag{66}$$

Note that if we choose $\theta = 0$, we get $\phi_1(c,\theta) = H(c)$ and $c_{\max} = 1/2$, which corresponds to our results for the q-exponential model. This means that in the large N limit, the q-exponential model behaves the same as the (1-exponential) ER model with $\theta = 0$ or equivalently with a link probability $p = 1/2$.

2. *Scaled q*

As shown in our previous analysis, the entire large-N behavior of the q-exponential model is encoded in the function $\phi_q(c,\theta,N)$ appearing in Eq. (56), as a consequence of the concentration of measure. We have just shown that, in the large N limit, $\phi_q(c,\theta,N)$ approaches a trivial limit regardless of q, if the latter remains fixed as N grows. We now show that this degeneracy disappears if, assuming the 'dense' regime where θ remains finite in the thermodynamic limit, we let q scale with N as

$$q \;=\; 1 \;-\; \frac{r}{N}\,, \tag{67}$$

where r is a finite constant (independent of N), which implies the finite limit

$$\lim_{N\to\infty}(1-q)N \;=\; r\,. \tag{68}$$

With this scaling, the second term in Eq.(56) remains finite and provides a new important contribution, genuinely inherent to the q-exponential model. We also see from Eq. (51) that the chosen scaling for q implies that all higher-order terms in the effective Hamiltonian become of the same order $\mathcal{O}(N)$ as $H(G,\theta)$, so that the q-ER model will take into account all possible k-link couplings (with $k = 1,\ldots,\infty$), each of them contributing with the same relevance.

To pave the way for the investigation of the effects of the scaling in Eq. (67), we note again that, for large N, the probability distribution becomes sharply peaked around the configurations that maximize the effective free entropy density $\phi_q(c,\theta,N)$. In this regime, we can carry out a *saddle-point approximation* and replace the ensemble average with the most likely value. With the reparametrization in Eq. (67), the effective free entropy density takes the form

$$\phi_r(c,\theta) \;=\; H_b(c) \;-\; \frac{1}{r}\ln(1+r\theta c)\,, \tag{69}$$

which reduces to its Shannonian counterpart in Eq. (64) as $r \to 0$. To find the maxima, we compute the derivative of $\phi_r(c,\theta)$ with respect to c, which reads

$$\frac{\partial}{\partial c}\phi_r(c,\theta) \;=\; \ln\left(\frac{1-c}{c}\right) \;-\; \frac{\theta}{1+r\theta c}\,. \tag{70}$$

Setting this derivative to zero leads to the *saddle-point equation*

$$c \;=\; \frac{1}{1\;+\;\exp\left(\frac{\theta}{1+r\theta c}\right)}\,. \tag{71}$$

We have seen in Eq. (29) that in the Shannonian case ($r = 0$), there is a single solution at $c = 1/(1+e^{\theta})$. For $r \neq 0$, however, the behavior can be different, as we show in the next section.

C. Phase transition

The saddle-point equation (71) allows us to characterize the effects of the link dependencies implied by the chosen scaling for q, showing that they lead to a phase transition similar to the one observed for traditional ERGs with higher-order constraints, even if the q-ER still enforces only a single constraint on the total number of nodes.

Indeed, Fig. 1 shows the effective free entropy density $\phi_r(c,\theta)$ for $r = 2.7$ at $\theta = 4, 5, 5.5$. For $\theta = 4$, a unique maximum is observed, whereas for $\theta = 5$ and $\theta = 5.5$ two local maxima emerge. Furthermore, the global maximum shifts abruptly between $\theta = 5$ and $\theta = 5.5$, suggesting the occurrence of a phase transition.

To confirm whether a phase transition occurs, we compute the local maxima and identify the global maximum as a function of θ. In Fig. 2 (right), we show the locations of the local maxima as a function of θ. For $r = 2.7$, two distinct local maxima, denoted c_{low} and c_{high}, exist within the interval $\theta \in [4.831, 5.846]$, corresponding

to a stable and a metastable state, respectively. The points $\theta_{sp-} \approx 4.831$ and $\theta_{sp+} \approx 5.846$ are the *spinodal points*. The phase transition occurs at $\theta_{\rm crit} \approx 5.154$, below which the high-density state is stable and above which the low-density state becomes stable. For comparison, Fig. 2 (left) shows the case $r = 2$, where no phase transition is observed. The Shannonian network density, given by $c_{\rm Shannon} = 1/(1+e^{\theta})$, is also shown for reference.

The conditions for the phase transition can be made precise. In Appendix G, we establish the following result:

Proposition 1. *Let $x^* > 0$ denote a solution of the equation*

$$\tanh\left(\frac{1}{x^*}\right) = x^* . \tag{72}$$

The phase transition in the q-ER model occurs if and only if

$$r \equiv (1-q)N > r_{\rm cp} \equiv \frac{(x^*)^2}{1-(x^*)^2} \approx 2.2768 . \tag{73}$$

The critical value $r_{\rm cp} \approx 2.2768$ is thus the threshold below which no phase transition can occur. At $r = r_{\rm cp}$, the free entropy density $\phi_{r_{\rm cp}}(c,\theta)$ first develops an inflection point as θ is varied; we define $\theta_{\rm cp}$ as this threshold value.

In Appendix H, we establish the following proposition:

Proposition 2. *The critical point $\theta_{\rm cp}$ is given by*

$$\theta_{\rm cp} = 2 + \sqrt{\frac{4(r_{cp}+1)}{r_{\rm cp}}} \approx 4.3994 . \tag{74}$$

We map out the full phase diagram by numerically solving the saddle-point equation (71) for a range of values of r. For each value we determine the two spinodal points $\theta_{\rm sp-}$ and $\theta_{\rm sp+}$, the phase-transition boundary $\theta_{\rm crit}$, and the critical point $(r_{\rm cp}, \theta_{\rm cp}) \approx (2.2767, 4.3994)$. The result is presented in Fig. 3.

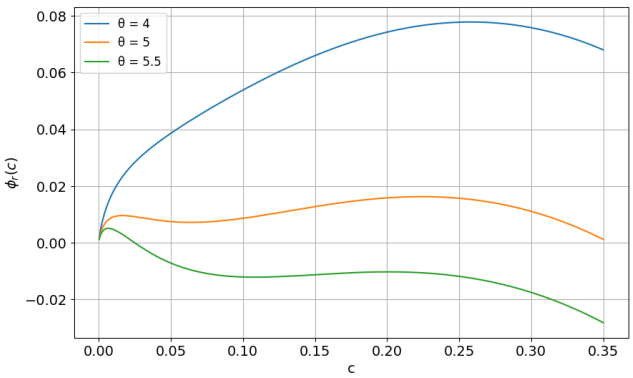

FIG. 1: The effective free entropy density $\phi_r(c,\theta)$ for $r = 2.7$ as a function of link density c for different values of θ. The global maxima of these curves determine the typical configuration.

Finally, we note that the presence of a phase transition has direct implications for practical applications, such as network reconstruction. In particular, for a given value of r there may be no choice of θ that reproduces a prescribed target network density. In the metastable region, two distinct phases coexist — a low-density phase $c_{\rm low}$ and a high-density phase $c_{\rm high}$. Consequently, densities lying between maximum value of $c_{\rm low}$ and minimal value of $c_{\rm high}$ cannot be reconstructed. This behavior is already evident in Fig. 2. For a range of r-values, we numerically determined these boundaries and identified the region in which density reconstruction is impossible. The results are shown in Fig. 4. We further assume that the irreconstructible region increases with r, and in the large-r limit, only allows either $c = 0$ or $c = 0.5$.

Remarkably, we have therefore found behavior typical of standard ERGs defined via the addition of higher-order constraints besides the overall number of links [2, 10, 24, 68–70]. However, in our q-ERG setting that behaviour does not actually require such extra constraints in the definition of the Hamiltonian. It is the regime $q \neq 1$ that automatically produces the phenomenology typical of higher-order ERGs, while keeping the overall link density as the only target constraint.

1. *Note on the order of limits*

In statistical physics, phase transitions are typically characterized by the non-analytic behavior of thermodynamic potentials, such as the Helmholtz or Gibbs free energy. Such non-analyticities, however, arise only in the thermodynamic limit $N \to \infty$. Consider again Eq. (55): indeed, the non-analyticity is caused by the maximization over c in the limit $N \to \infty$. However, we previously argued that in this limit the model approaches the Shannonian regime with $\theta = 0$, for which no phase transition is present.

This apparent contradiction arises from the fact that we imposed $r = (1-q)N$ and treated r as fixed. Doing so implicitly changes the structure of the limiting procedures. If we first take the limit $N \to \infty$ at fixed q, we indeed recover the degenerate Shannonian model in which the influence of the parameter θ vanishes, as previously discussed. Taking the subsequent limit $q \to 1$ does not alter this conclusion. Conversely, if we instead take the limit $q \to 1$ first, we recover the standard Erdős–Rényi model. This *non-commutativity of limits* was already emphasized in [72], where it was shown that Tsallis entropy lie in the $(c,d) = (q,0)$ universality class, whereas Shannon and Rényi entropies correspond to $(c,d) = (1,1)$. These belong to different scaling classes and therefore exhibit distinct behaviors in the thermodynamic limit $N \to \infty$. The situation is also reminiscent of the standard theory of phase transitions, where

$$\lim_{N\to\infty} \lim_{T\to T_c} \neq \lim_{T\to T_c} \lim_{N\to\infty} , \tag{75}$$

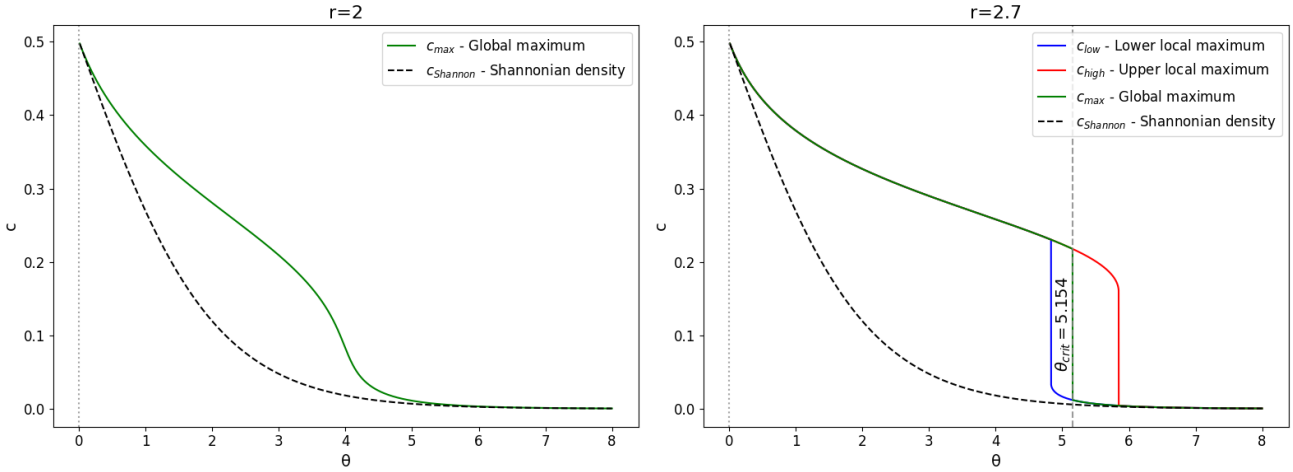

FIG. 2: On the left, we show the behavior of the typical configuration link density c_{max} for $r = 2$ as a function of θ. On the right, for $r = 2.7$, we plot the local and global maxima of the effective free entropy density $\phi_r(c, \theta)$ as a function of θ. In the range $\theta \in [4.831, 5.846]$, there are two local maxima c_{low} and c_{high}, where one is stable and one metastable. The phase transition happens at $\theta_{crit} \approx 5.154$. For reference, we also show the Shannonian network density $c_{Shannon} = (1 + e^{\theta})^{-1}$.

with T_c denoting the critical temperature of the phase transition.

Our central observation is that the $N \to \infty$ limit taken with $(1 - q)N$ held constant yields nontrivial behavior, and that a phase transition emerges only under this scaling. Moreover, we observe signatures of this transition to persist even at finite N, manifesting, for example, through metastable states. In network ensembles, where N (the number of possible links) is typically large but fixed, the saddle-point approximation should therefore remain applicable.

Let us now examine the implications of the scaling $r = (1-q)N$ for the entropy. In particular, the Tsallis entropy may be equivalently rewritten in terms of the parameter r as

$$T_r[P] = \frac{N}{r}\left(1 - \sum_{G \in \mathcal{G}_n} [P(G)]^{r/N+1}\right). \tag{76}$$

By considering the uniform distribution $P_u(G) = 1/W$ for all $G \in \mathcal{G}_n$, where $W = 2^N$ is the number of possible binary undirected graphs with n nodes and $N = n(n-1)/2$ node pairs, we obtain

$$T_r[P_u] = \frac{N}{r}\left(1 - W^{-r/N}\right) = N\frac{1 - 2^{-r}}{r} \propto N, \tag{77}$$

meaning that the entropy is extensive! So, by the specific choice of the limiting scaling $\lim_{N\to\infty}(1 - q)N = r$ with

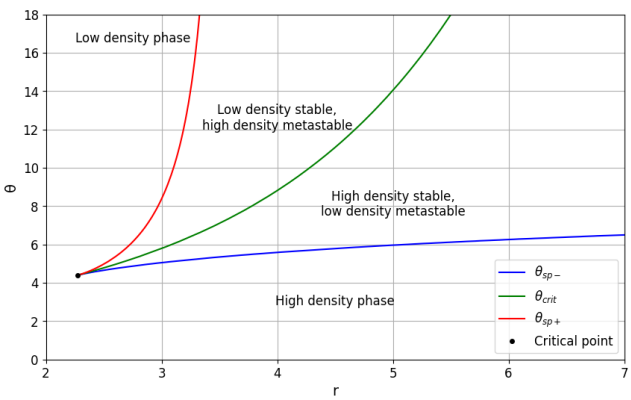

FIG. 3: The phase diagram of the q-ER model. For each value of r, we show the spinodal points $\theta_{\mathrm{sp}-}$ and $\theta_{\mathrm{sp}+}$ as well as the phase transition θ_{crit}. The critical point $(r_{\mathrm{cp}}, \theta_{\mathrm{cp}}) \approx (2.2767, 4.3994)$ is shown as the black dot.

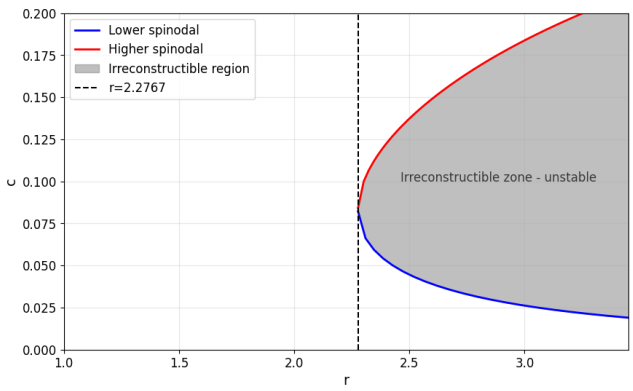

FIG. 4: For different values of r, we numerically approximated boundaries of the region, where density reconstruction in the q-ER model is not possible. Such a region is determined by the lowest possible value of c_{high} and the highest possible value of c_{low} in the metastable regime. We also show r value of the critical point, $r_{\mathrm{cp}} = 2.2767$.

r finite, we actually made the Tsallis entropy extensive again. However, remarkably, the resulting behavior is much richer than in the Shannonian case, which (for a given exponential state-space scaling) is extensive as well.

D. Moments and cumulants

The saddle-point approximation identifies the typical link density in the large-N limit, but does not capture fluctuations or correlations between links. To access these, we now compute the exact moments and cumulants under the q-exponential model, starting with the average number of links. In Shannonian case we know that

$$\langle L\rangle \;=\; -\frac{\partial \ln \mathcal{Z}(\theta)}{\partial\theta}\,. \tag{78}$$

This equation, however, does not hold in the q-exponential case. In particular, in Appendix C, we show that instead we have

$$\langle L\rangle \;=\; N\exp_{2-q}(-\theta)\frac{\mathcal{Z}_{q,N-1}\left(\frac{\theta}{1-(q-1)\theta}\right)}{\mathcal{Z}_{q,N}(\theta)}\,. \tag{79}$$

Note, in particular, the appearance of

$$\mathcal{Z}_{q,N-1}\left(\frac{\theta}{1-(q-1)\theta}\right)\,, \tag{80}$$

i.e., the partition function for a system with one fewer link in the sum, and with a modified parameter. Thus, the general relationship is not expressed via a derivative, but through an explicit formula connecting partition functions with a different number of summed links. It might be easily checked that, as $q\to 1$, Eq. (79) reduces to Eq. (78).

The marginal probability of a single link is, due to permutation symmetry of the probability, just $\langle a_{ij}\rangle = \langle L\rangle/N = \langle c\rangle$ for all (i,j), i.e. it is the same as the average link density. We therefore have

$$\langle a_{ij}\rangle \;=\; \langle c\rangle \;=\; \exp_{2-q}(-\theta)\frac{\mathcal{Z}_{q,N-1}\left(\frac{\theta}{1-(q-1)\theta}\right)}{\mathcal{Z}_{q,N}(\theta)}\,. \tag{81}$$

In Appendix D, we show that, instead of computing $\langle a_{ij}\rangle$ using $\langle L\rangle$, there is actually an elegant way to compute it directly. This is because one can easily sum over other degrees of freedom. The same methodology can be extended to the computation of higher-order joint moments. For k distinct links, the following identity is derived in Appendix E:

$$\langle a_{i_1j_1}\cdots a_{i_kj_k}\rangle = \exp_{2-q}(-k\theta)\frac{\mathcal{Z}_{q,N-k}\left(\frac{\theta}{1-k(q-1)\theta}\right)}{\mathcal{Z}_{q,N}(\theta)}\,. \tag{82}$$

Although the above expression is elegant, it presents a computational challenge due to the difficulty of evaluating the partition function. Its characteristic asymptotic behavior is known, see Eq. (59), however, its exact value remains unknown, even for large N. Since, as discussed above, a particularly relevant regime corresponds to values of q close to one, we now turn to a perturbative analysis of this case.

In terms of the effective Hamiltonian $H_q^{\rm eff}(G,\theta)$ introduced in Eq. (51), the partition function is given by

$$\mathcal{Z}_{q,N} \;=\; \sum_{G\in\mathcal{G}_N}\exp\!\big(-H_q^{\rm eff}(G,\theta)\big)\,. \tag{83}$$

By introducing source terms h_{ij} coupled to individual links, we can then derive the standard identities

$$\mathcal{Z}_{q,N}(\theta,\{h_{ij}\}) \;=\; \sum_{G\in\mathcal{G}_N}\exp\Big(-H_q^{\rm eff}(G)-\sum_{i=1}^{n}\sum_{j<i}h_{ij}a_{ij}\Big)\,, \tag{84}$$

and

$$\left.\frac{\partial\ln\mathcal{Z}_{q,N}}{\partial h_{kl}}\right|_{\{h_{ij}\}=0} \;=\; \langle a_{kl}\rangle\,, \tag{85}$$

$$\left.\frac{\partial^2\ln\mathcal{Z}_{q,N}}{\partial h_{kl}\partial h_{uv}}\right|_{\{h_{ij}\}=0} \;=\; \langle a_{kl}a_{uv}\rangle \;-\; \langle a_{kl}\rangle\langle a_{uv}\rangle\,, \tag{86}$$

the last expression coinciding with the covariance $\mathrm{Cov}(a_{kl},a_{uv})$. By isolating the Shannon-entropy contribution from the effective Hamiltonian, we can write

$$\begin{aligned}\mathcal{Z}_{q,N}(\theta,\{h_{ij}\}) \;&=\; \sum_{\{a_{ij}\}}\exp\Big[-\sum_{i=1}^{n}\sum_{j<i}(\theta-h_{ij})a_{ij}\Big]\\ &\quad\times\;\exp(-\Delta H_q(G))\\ &=\; \mathcal{Z}_{1,N}(\theta,\{h_{ij}\})\;\langle\exp(-\Delta H_q)\rangle_{\rm Sh}\,,\end{aligned} \tag{87}$$

where $\langle\cdots\rangle_{\rm Sh}$ denotes the ensemble average with respect to the Shannonian distribution, and $\Delta H_q(G)$ denotes the non-Shannonian part of the effective Hamiltonian, which can be expanded as

$$\begin{aligned}\Delta H_q(G) \;&=\; \tfrac{1}{2}(q-1)\theta^2L(G)^2 \;+\; \tfrac{1}{3}(q-1)^2\theta^3L(G)^3\\ &\quad+\; \tfrac{1}{4}(q-1)^3\theta^4L(G)^4 \;+\; \mathcal{O}\big((q-1)^4\big)\,.\end{aligned} \tag{88}$$

To compute the cumulants, we need the logarithm of the partition function, i.e.

$$\begin{aligned}\ln\mathcal{Z}_{q,N}(\theta,\{h_{ij}\}) \;&=\; \ln\mathcal{Z}_{1,N}(\theta,\{h_{ij}\})\\ &\quad+\; \ln\langle\exp(-\Delta H_q)\rangle_{\rm Sh}\\ &=\; \prod_{i=1}^{n}\prod_{j<i}\ln(1+e^{-\theta+h_{ij}})\\ &\quad+\; \ln\langle\exp(-\Delta H_q)\rangle_{\rm Sh}\,.\end{aligned} \tag{89}$$

We can then use the cumulant expansion

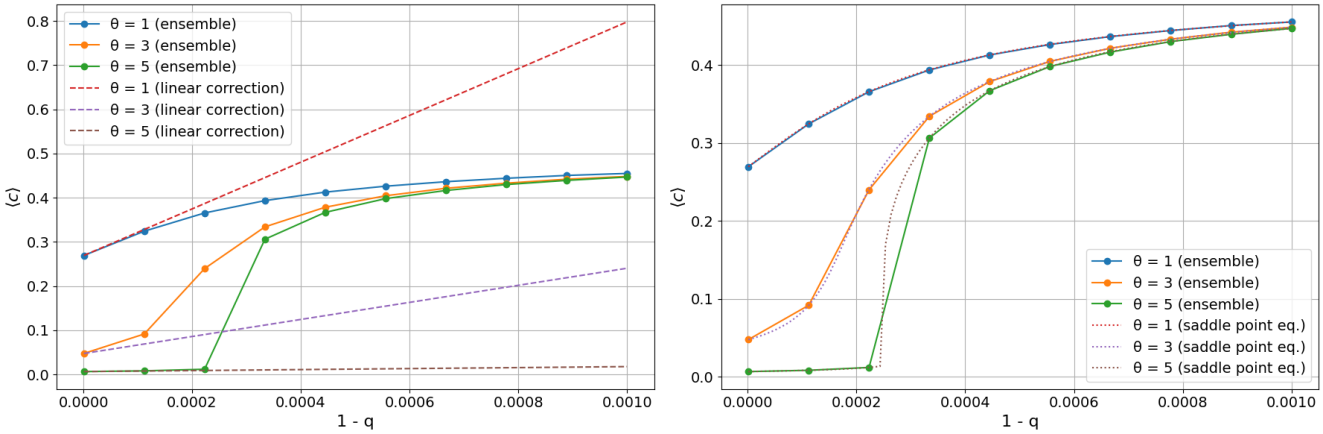

FIG. 5: For ensembles of 1000 graphs with 100 nodes, we compute the average link density $\langle c\rangle$ and compare with the perturbative linear correction [Eq. (91)] and with the solutions of the saddle-point equation [Eq. (71), on the right].

$$\begin{aligned}\ln\langle\exp(-\Delta H_q)\rangle_{\rm Sh} &= -\langle\Delta H_q\rangle_{\rm Sh} + \frac{1}{2}\Big(\langle\Delta H_q^2\rangle_{\rm Sh} - \langle\Delta H_q\rangle_{\rm Sh}^2\Big) + \mathcal{O}\big((q-1)^3\big) \\ &= -\frac{1}{2}(q-1)\theta^2\langle L^2\rangle_{\rm Sh} + (q-1)^2\left[-\frac{1}{3}\theta^3\langle L^3\rangle_{\rm Sh} + \frac{1}{8}\theta^4\left(\langle L^4\rangle_{\rm Sh} - \langle L^2\rangle_{\rm Sh}^2\right)\right] + \mathcal{O}\big((q-1)^3\big)\,. \quad (90)\end{aligned}$$

To extract the cumulants, we differentiate $\ln \mathcal{Z}_{q,N}$ with respect to the source terms h_{ij} as prescribed by (85) and (86). These computations are carried out in Appendix F, yielding, to first order in $(q-1)$:

$$\langle a_{ij}\rangle = \frac{1}{1+e^\theta}\left[1 + \frac{1}{2}(1-q)\theta^2\frac{e^\theta}{(1+e^\theta)^2}(2N-1+e^\theta)\right] + \mathcal{O}((q-1)^2)\,, \quad (91)$$

$$\mathrm{Cov}(a_{ij},a_{kl}) = (1-q)\frac{\theta^2 e^{2\theta}}{(1+e^\theta)^4} + \mathcal{O}((q-1)^2)\,. \quad (92)$$

As expected from the higher-order link interactions induced by $q \neq 1$, the covariance is positive for $q < 1$ and vanishes at leading order as $q \to 1$. Notably, to first order in $(q-1)$ it is independent of N, though this independence breaks down at higher orders.

E. Network properties

Having established the theoretical properties of the q-ER model defined by Eq. (49), we now turn to numerical experiments, both to validate our analytical predictions and to study further network properties such as clustering and average nearest-neighbor degree. Graph ensembles are sampled using the Metropolis–Hastings algorithm with a symmetric link-flip proposal (see Appendix K), for which the acceptance probability takes the form

$$\alpha = \min\left(1, \frac{P_q(G')}{P_q(G_t)}\right) = \min\left(1, \frac{\exp_{2-q}(-\theta L(G'))}{\exp_{2-q}(-\theta L(G_t))}\right), \quad (93)$$

where the partition function cancels in the ratio.

1. *Link density and percolation*

We considered two complementary methods for approximating the link density. The first one is a perturbative approach based on Eq. (91), while the second one involved solving the saddle-point equation (71). Fig. 5 compares the predictions of these approaches with numerical simulations. For 10 values of $(1-q)$ evenly spaced between 10^{-6} and 10^{-3}, and for $\theta \in \{1,3,5\}$, we generated ensembles of 1000 graphs with 100 nodes, computed their link densities, and compared them with the theoretical predictions. The results indicate that the linear perturbative approximation loses accuracy rapidly, whereas the saddle-point solution remains in excellent agreement with the numerical data. Furthermore, as q deviates further from unity, the link density for all values of θ approaches 0.5, consistent with our analytical predictions.

Let us next examine, how the link densities are

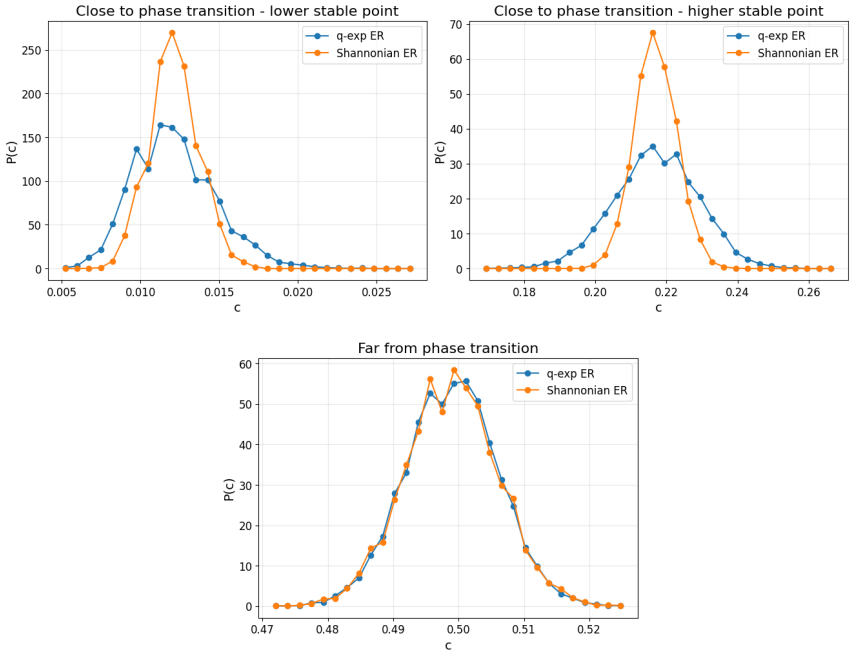

FIG. 6: Comparison of distributions link densities for two situations: close to the phase transition (top two graphs) with $r = 2.7$ and $\theta = 5.154$ and far from it (bottom graph) with $r = 1000$ and $\theta = 1$. We also plot the corresponding Erdős–Rényi ensembles (fitted to reproduce the same average link density). For the top two graphs, we used two different initial conditions and converged to two different fixed points of the saddle-point equation. Used ensembles contained 10000 graphs with 100 nodes.

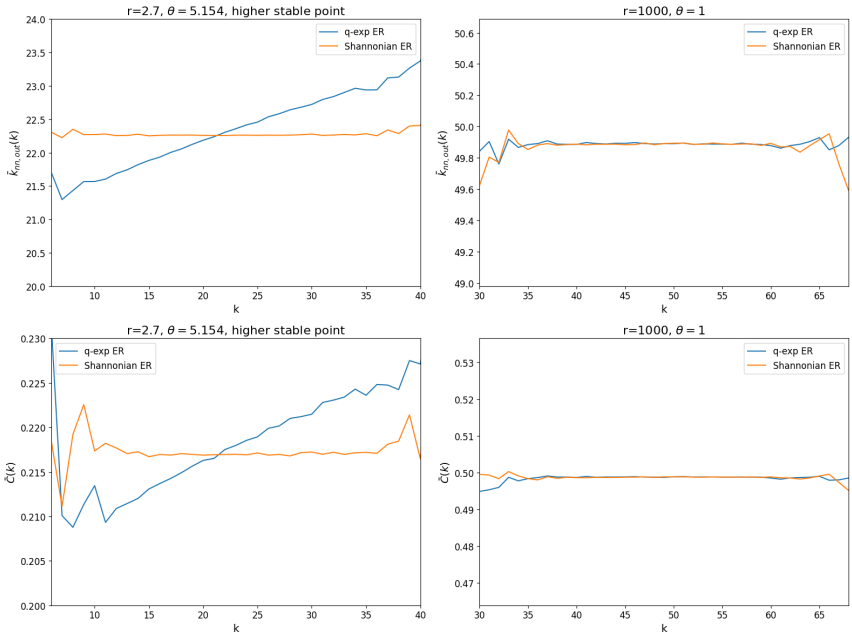

FIG. 7: Comparison of the ANND and clustering coefficient depending on node degree for the non-Shannon ensembles close to the phase transition (left) and far from it (right). Interestingly, near the phase transition, the q-exponential model develops positive assortativity and increasing clustering with degree. For extreme values of degrees, there are less available data and therefore the computation is more noisy. Here we have used the same ensembles as for Fig. 6.

distributed among the ensemble. Two representative regimes are considered: one close to the phase transition and one far from it. We focus on the case $r = (1-q)N = 2.7$, for which a phase transition was identi-

fied at $\theta_{\rm crit} = 5.154$, with the corresponding saddle-point solutions $c_{\rm low} = 0.0118$ and $c_{\rm high} = 0.2177$. To probe the resulting distributions, we generate two ensembles, each with a different initial condition for sampling: one starting with a graph with density $c_{\rm init} = 0.01$ and the other with $c_{\rm init} = 0.3$. For each ensemble, 10^4 graphs are generated, and the resulting link-density histograms are computed. For comparison, we also generate an ensemble with $r = 1000$ and $\theta = 1$, corresponding to a regime far from the transition. Finally, we construct Shannon ensembles tuned to reproduce the same average link densities as the corresponding non-Shannon ensembles.

The results are summarized in Fig. 6. The two distinct initial conditions converge to different fixed points of the saddle-point equation, as we can see from the corresponding link-density distributions. The measured mean values, $\langle c\rangle_{\rm low} = 0.0120$ and $\langle c\rangle_{\rm high} = 0.2173$, are in excellent agreement with the theoretical predictions. Notably, in the vicinity of the phase transition, the non-Shannon ensembles exhibit substantially broader distributions than their Shannon counterparts. The corresponding standard deviations are $\sigma_{\rm low} = 0.0018$ and $\sigma_{\rm high} = 0.0082$ for the non-Shannon ensembles, compared to $\sigma_{\rm low} = 0.0011$ and $\sigma_{\rm high} = 0.0041$ for the Shannon ensembles.

By contrast, for the ensemble far from the critical point, the distributions are nearly indistinguishable, with $\langle c\rangle = 0.4995$ and $\sigma = 0.005$ in both cases. This behavior confirms our analytical expectation that, for q sufficiently far from unity, the model reduces to the Erdős–Rényi random graph with link probability 1/2, whereas for q close to unity the two ensembles display markedly different statistical properties.

We also investigated the percolation transition, which is a characteristic phenomenon of the ER model. The overall behavior appears to be the same for the q-ER, the only noticeable difference being an increased variance in the size of the giant connected component. Further results are provided in Appendix I.

2. *Average nearest-neighbor degree and clustering*

Let us now compare additional structural properties of the networks. We find no significant differences in the degree distributions between the q-exponential and exponential (Shannon) ensembles. In contrast, the average nearest-neighbor degree (ANND) and the clustering coefficient exhibit notable differences between the q-exponential model near the phase transition and the regime far from criticality. These results are shown in Fig. 7, where we analyze the same ensembles as in Fig. 6 and evaluate the degree dependence of the ANND and clustering coefficient.[4]

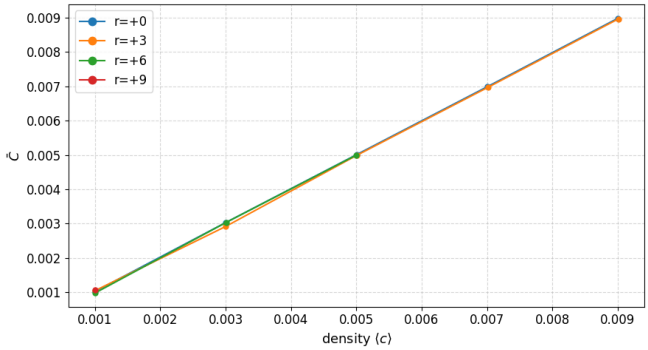

FIG. 8: Average clustering coefficient as a function of graph density for the q-ER model, for target average degrees $\langle k\rangle \in \{1,3,5,7,9\}$ and $r \in \{0,3,6,9\}$, based on ensembles of 1000 graphs with $n = 1000$ nodes. No significant deviation from the Shannon ($r = 0$) case is observed, in contrast to the q-exponential CM (cf. Fig. 10).

For the ANND, the non-Shannon model near the transition displays positive assortativity, accompanied by an increase of the clustering coefficient with node degree. We attribute this behavior to the presence of positive edge-to-edge correlations, in agreement with the analytical prediction in Eq. (92). By contrast, in the regime $r = 1000$ and $\theta = 1$, the model approaches the Shannon limit, and both assortativity and clustering remain approximately constant, as expected for the Erdős–Rényi model.

We also studied the overall average clustering coefficient as a function of graph density for varying r, shown in Fig. 8. Across all target degrees $\langle k\rangle \in \{1,3,5,7,9\}$ and values of $r \in \{0,3,6,9\}$, the clustering remains essentially indistinguishable from the Shannon ($r = 0$) baseline. We attribute this to the nature of the higher-order interactions in the q-ER model: as seen in Eq. (51), the corrections couple all pairs of links equally through powers of the total link count $L(G)$, with no separate parameter to enforce a preference for triangle formation. This is in sharp contrast to what we will show in the next Section for the q-exponential CM, where varying r produces substantial deviations from the Shannon case and has a remarkable impact on the clustering coefficient.

V. q-EXPONENTIAL CONFIGURATION MODEL

We finally turn to the q-exponential generalization of the CM introduced in Sec. II B (which we denote as the q-CM), which has constraints on the degrees of all nodes. The corresponding Hamiltonian is still given by

[4] For each graph, the ANND and clustering coefficient are first computed at the node level. Nodes with the same degree are then grouped across the ensemble, and the corresponding quantities are averaged within each group.

Eq. (33), while the resulting q-exponential graph probability is given by

$$P_q(G) = \frac{1}{\mathcal{Z}_q(\boldsymbol{\theta})} \exp_{2-q}\left(-\sum_{i=1}^{n}\sum_{j<i}(\theta_i+\theta_j)a_{ij}\right), \quad (94)$$

with partition function

$$\mathcal{Z}_q(\boldsymbol{\theta}) = \sum_{\{a_{ij}\}} \exp_{2-q}\left(-\sum_{i=1}^{n}\sum_{j<i}(\theta_i+\theta_j)a_{ij}\right). \quad (95)$$

For given q, the conditions for enforcing a target degree sequence $\{k_i^*\}_{i=1}^n$ are analogous to those in Eq. (37):

$$k_i^* \equiv \langle k_i\rangle = \sum_{j\neq i}\langle a_{ij}\rangle = \sum_{j\neq i} p_{ij}, \quad i=1,n, \quad (96)$$

where here $p_{ij} = \langle a_{ij}\rangle$ represents the *marginal* connection probability and, however, the overall $P_q(G)$ does not factorize into a product of terms involving p_{ij}, because of overall dependencies between all pairs of nodes, as for the q-ER model.

We can compute $\langle a_{kl}\rangle$ for the q-CM by summing out all degrees of freedom except link (k,l) (see Appendix L for the derivation). Introducing the reduced partition function and reparametrized multipliers

$$\tilde{\mathcal{Z}}_q^{(k,l)}(\boldsymbol{\theta}) \equiv \sum_{\substack{\{a_{ij}\}\\(i,j)\neq(k,l)}} \exp_{2-q}\left(-\sum_{\substack{i<j\\(i,j)\neq(k,l)}}(\theta_i+\theta_j)a_{ij}\right),$$

$$\tilde{\theta}_i^{(k,l)} \equiv \frac{\theta_i}{1-(q-1)(\theta_k+\theta_l)},$$

we obtain

$$\langle a_{kl}\rangle = \exp_{2-q}\left(-(\theta_k+\theta_l)\right)\frac{\tilde{\mathcal{Z}}_q^{(k,l)}(\tilde{\boldsymbol{\theta}}^{(k,l)})}{\mathcal{Z}_q(\boldsymbol{\theta})}. \quad (97)$$

This expression is reminiscent of Eq. (81) obtained in the q-ER model. Although it would be desirable to further simplify the numerator, $\tilde{\mathcal{Z}}_q^{(k,l)}(\tilde{\boldsymbol{\theta}}^{(k,l)})$, the q-exponential of a sum does not factorize in a simple manner, and Eq. (97) is likely the most compact form attainable. In principle, we would again carry out a perturbative analysis, analogous to the one performed in Sec. IV D. In what follows, however, we focus primarily on the behavior of network properties and evaluate them numerically.

A. Network properties

Among the virtually infinite possible choices for generating target degree sequences, we consider the convenient strategy indicated in Eqs. (38) and (39): we sample $\{s_i\}_{i=1}^n$ i.i.d. from a given distribution, while adjusting z to enforce a desired overall link density c^*. To pick a realistic distribution, we refer to a study of interbank networks [73] where the empirical distribution of node 'strengths' $\{s_i\}_{i=1}^n$ was found to be well approximated by a log-normal distribution with $\sigma = 2.28$, $\mu = -\sigma^2/2$. Since the relationship to the parameters θ_i is given by $x_i = e^{-\theta_i}$, we can sample θ_i from a normal distribution with $\sigma = 2.28$ and $\mu = \sigma^2/2$. Inserting these values into the p_{ij} in Eq. (38) and summing over all pairs as in Eq. (39), we obtain the implied link density in the reference ($q=1$) Shannonian CM and we adjust z to enforce a desired value c^* that we consider as the target one. We then consider different values of q (or r) and recalculate z for each of them, to match the same target density c^* for the general q-CM, while keeping strengths $\{s_i\}_{i=1}^n$ the same. With this strategy we can vary z to explore the space of densities, and compare the CM and q-CM for the same target level of density. Since the parameter z is multiplicative, in terms of θ_i we have to find an additive shift denoted by θ_{shift}, which will recover the target density of the network.

Let us now see whether the q-exponential model gives significant corrections in terms of network properties and whether we will also observe signs of a phase transition. To this end, we choose target average node degrees $k^* \in \{1,3,5,7,9\}$ (which corresponds to choosing target average graph density $c^* = k^*/(n-1)$. For each of them, we choose several values of $r = (1-q)N$ (N is the maximum number of links, $N = n(n-1)/2$) and for each we find the corresponding θ_{shift} to recover the target density[5]. Importantly, this time we try both $r > 0$ (corresponding to $q < 1$) and $r < 0$ (corresponding to $q > 1$). We then sample 1000 networks with $n = 1000$ nodes using the Metropolis–Hastings algorithm, similarly to the q-ER model (Sec. IV E), with a symmetric link-flip proposal (see Appendix K). The acceptance probability is given by

$$\alpha = \min\left(1, \frac{P_q(G')}{P_q(G_t)}\right) = \min\left(1, \frac{\exp_{2-q}\left(-\sum_{i=1}^n \theta_i k_i(G')\right)}{\exp_{2-q}\left(-\sum_{i=1}^n \theta_i k_i(G_t)\right)}\right). \quad (98)$$

We find that, similar to the q-ER model, certain target densities are irreconstructible, signifying a similar type of phase transition. For the values of r for which reconstruction was possible, however, we observe rich degree-resolved structure. Fig. 9 shows the complementary cumulative degree distribution (CCDF), average nearest-neighbor degree $\bar{k}_{nn}(k)$, and clustering coefficient $\bar{C}(k)$ for target average node degree $k^* = 3$ across a range of r values. A full empirical study across other target degrees

[5] We find the corresponding θ_{shift} by bisection, i.e. we choose an interval of possible values of θ_{shift} and iteratively narrow it down until we reach given precision.

is presented in Appendix J. The degree distributions are nearly indistinguishable across r, reflecting the fact that the θ_i distribution is held fixed and only the additive shift θ_{shift} is tuned to recover the target density. In contrast, both the ANND and the clustering coefficient are strongly sensitive to r: more negative values of r (corresponding to $q > 1$) produce markedly higher ANND and clustering at low degrees, while positive r suppresses them. We therefore find tunable clustering and assortativity profiles, similar to those reported in studies where the degree sequence is kept fixed and extra constraints on the clustering coefficients are introduced [74]. However, here the only actual constraint is the degree sequence alone as in the traditional CM, and the tunable clustering profiles are automatically produced by values $q \neq 1$.

A striking consequence is that the parameter r acts as an independent knob for the average local clustering coefficient *at fixed target density*, in sharp contrast to the standard (Shannon) CM where the degree sequence alone determines the clustering. Indeed, as shown in Fig. 10, there is a broad range of accessible values of the average clustering coefficient for a given target density, demonstrating that the q-exponential CM provides a genuinely richer structural repertoire than its Shannonian counterpart. Remarkably, this allows for the coexistence of sparsity and clustering, which is a widespread empirical property of real networks that is impossible to replicate in the usual $q = 1$ case, for both the ER and the CM. Moreover, comparing Fig. 10 with Fig. 8, we stress that the dependence, for fixed density, of the average clustering coefficient on the value of q is found only for the q-CM and not for the q-ER model. This means that the mechanism required to produce a non-trivial clustering is not the interaction among links *per se*, but the *combination of interaction and heterogeneity*. Indeed, as already recalled in Sec. IV A, also in traditional homogeneous ERGs such as the edge-triangle model [2, 10, 70] it is not possible to obtain simultaneously a large clustering and a low density (and more generally any combination of values for the two quantities), despite the presence of interactions and the existence of two separate Lagrange multipliers designed to tune the two quantities. On the other hand, heterogeneity alone (as in the ordinary CM) is not enough to produce a large clustering. Therefore the combination of the two ingredients seems necessary.

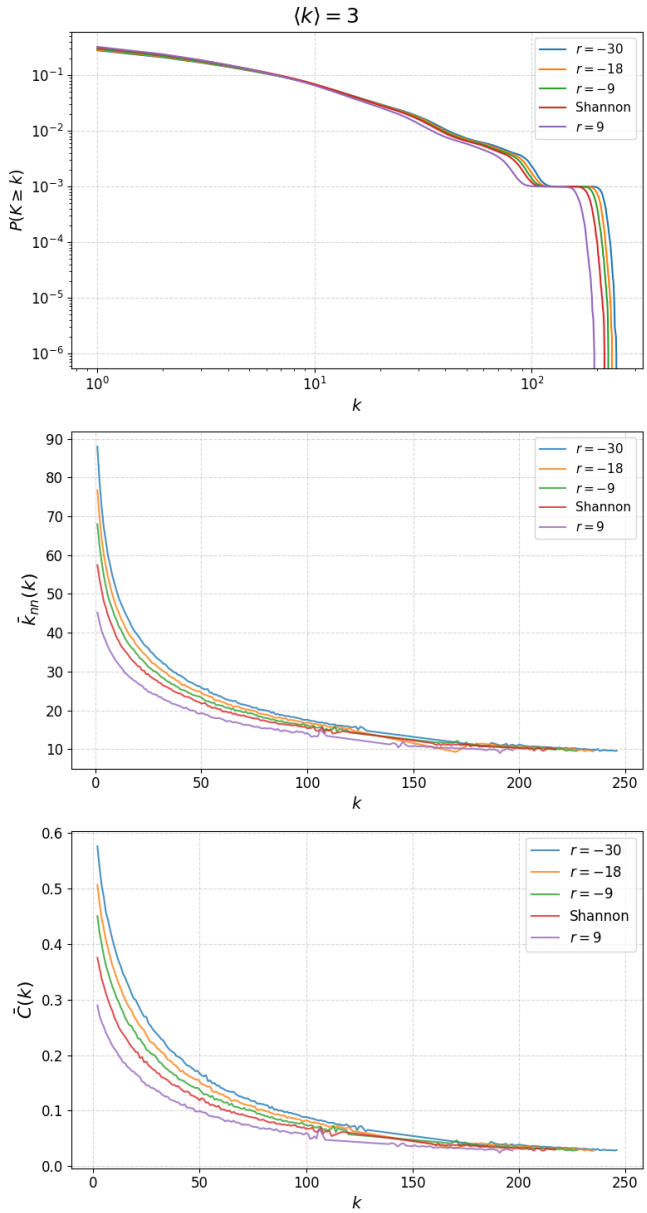

FIG. 9: Degree-resolved network properties of the q-exponential CM for target average degree $k^* = 3$ and $n = 1000$ nodes, for $r \in \{-30, -18, -9, 0, 9\}$. *Top*: complementary cumulative degree distribution. *Middle*: average nearest-neighbor degree $\bar{k}_{nn}(k)$. *Bottom*: clustering coefficient $\bar{C}(k)$. While the degree distributions are nearly identical across r values, both the ANND and clustering are strongly modulated by r, with more negative r producing higher correlations and clustering at low degrees. Each curve is averaged over an ensemble of 1000 graphs.

VI. CONCLUSIONS

In this paper, we have introduced the most agnostic inferential framework for network ensembles with given expected properties, based on the maximization of generalized (Uffink-type) entropies that extend beyond Shannon entropy. The generalization is required in light of the recent (re)discovery [43–45] that the axiomatic definition of a proper inferential MEP unavoidably leads to a class of entropies that is more general than Shannon's, unlike initially claimed [37, 38]. The corresponding maximum-entropy probability distribution is q-exponential, with $q = 1$ reducing to the traditional Shannon case. The q-exponential turns the entire class of ERGs, which include the most popular network models, into q-ERGs, which we have investigated here for the first time. The resulting framework provides

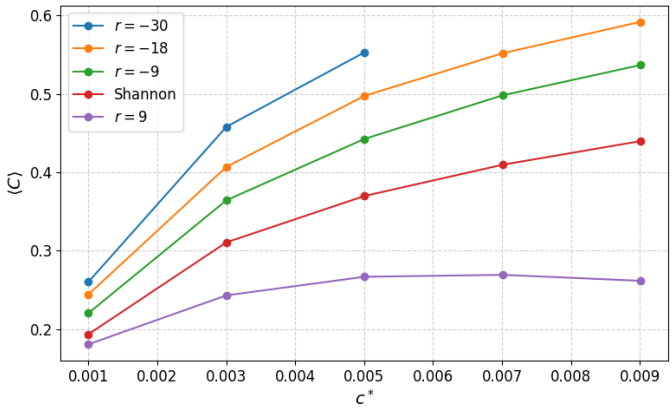

FIG. 10: Average clustering coefficient as a function of target network density c^* for the q-exponential CM, for target average node degrees $k^* \in \{1, 3, 5, 7, 9\}$. Each band of curves corresponds to a fixed target density with varying r, illustrating that the parameter r independently controls the clustering coefficient while leaving the degree sequence unchanged. The accessible range of clustering grows substantially with increasing target density. Each data point is averaged over an ensemble of 1000 graphs on $n = 1000$ nodes.

a promising link between the simplicity of maximum-entropy inference (within the MEP approach) and the complexity of real-world network topology. Within this setting, the structure of the underlying Uffink MEP functionals admits a clear operational meaning: the function f determines the scaling behavior of the entropy for uniform distributions (which is relevant for classification purposes [46, 75]), while the positive parameter q quantifies correlations in the system through the associated MEP distributions.

Our analysis of the simplest q-ERGs (defined by enforcing the overall link density and the degree sequence respectively) yielded several key insights:

- **Non-independence of links:** Unlike the classical Shannonian model, where links are independent Bernoulli variables, the q-exponential framework couples all potential links through the non-additive nature of the entropy. This can be efficiently captured via the notion of effective Hamiltonian, which features extra effective properties that will appear in the network unavoidably as $q \neq 1$, but without dedicated extra Lagrange multipliers tuning them: the extra features are all driven by the value of q, with no need to make arbitrary choices of extra constraints to be added to the original Hamiltonian. We derived perturbative expressions for covariances, showing that for $q \neq 1$, the model inherently induces nonzero correlations between links.

- **Higher-order networks from simple constraints:** The q-exponential setting provides a natural way of creating higher-order link dependencies from the 'usual' simple constraints (link density, degree sequence) without introducing such dependencies in an *ad hoc* manner via the addition of arbitrary higher-order constraints (such as the counts of two-stars, triangles, etc.). This result is parsimonious and particularly relevant in light of the current discussion around higher-order networks and hypergraphs [56–60]. We are therefore confronted with the fascinating result that, in absence of prior information ruling out values $q \neq 1$, the most agnostic maximum-entropy ensemble of graphs with given number of links, or given degree sequence, does not have independent edges and is therefore a higher-order network model, without the need to input additional (and arbitrarily chosen) higher-order constraints.

- **The critical scaling regime:** We identified that the 'Shannon degeneracy' (the tendency of q-exponential statistics to collapse into exponential Boltzmann–Gibbs statistics in the large-N limit) can be overcome by the specific scaling $(1-q)N = r$ with r fixed and finite. This reveals that for the q-exponential paradigm to remain relevant for large networks, the non-extensivity parameter $1-q$ must scale inversely with the system size.

- **Phase transitions and structural complexity:** Through a saddle-point analysis, in the q-ER model we proved the existence of a phase transition between sparse and dense network regimes. Crucially, we observed that near this criticality, the ensemble exhibits structural features — such as degree-dependent clustering and assortativity — that are absent in the standard Erdős–Rényi model. The behavior of the q-ER is in fact typical of standard homogeneous ERGs with the addition of higher-order constraints [2, 10, 68–70], which are however unnecessary here.

- **Generalized Configuration Models:** Our extension of the CM to the q-CM demonstrates that the q-exponential framework is sufficiently robust to accommodate degree constraints. We showed that the parameter r enables the tuning of higher-order structural properties, such as the average nearest-neighbor degree and the local clustering coefficient, while keeping the degree sequence constrained, thereby providing an additional degree of freedom for network reconstruction. The resulting tunable clustering and assortativity profiles are similar to those obtained in more complicated models [74] requiring the introduction of extra constraints on the local clustering coefficients. Remarkably, for appropriate values of q the average local clustering coefficient can be made large even for low link density, a feature that is widespread in real networks but impossible to replicate via the ordinary CM, via the q-ER, or even via traditional

homogeneous ERGs with edge-triangle constraints. This indicates the need for a combination of link dependencies and heterogeneity in order to replicate sparse clustered graphs via maximum-entropy models.

In future work, we plan to extend this framework to other class of graphs such as multiplex and weighted networks, where the non-additive nature of the entropy may additionally capture inter-layer correlations and the heavy-tailed nature of various distributions commonly observed in empirical data. Overall, our results indicate that the 'physics of information' in networks is highly sensitive to the choice of entropic functional, and that non-Shannon entropies offer a mathematically rigorous and phenomenologically viable approach for modeling the intrinsic dependencies within complex systems.

Acknowledgments

We thank Constantino Tsallis and Andrea Somazzi for useful discussions and comments on the manuscript. P.J. was supported by the Czech Science Foundation Grant (GAČR), Grant No. 25-18105S. This publication is part of the project "Redefining renormalization for complex networks" with file number OCENW.M.24.039 of the research programme Open Competition Domain Science Package 24-1, which is (partly) financed by the Dutch Research Council (NWO) under the grant https://doi.org/10.61686/PBSEC42210.

Appendix A: Derivation of the q-exponential ERG distribution

We derive the general q-exponential graph probability stated in Sec. III, starting from the Lagrange functional $\mathcal{L}_q[P]$ in Eq. (41). Computing the partial derivative with respect to $P_q(G')$ and demanding it to vanish for all $G' \in \mathcal{G}_n$, we find

$$\frac{\partial}{\partial P_q(G')}\mathcal{L}_q[P] \;=\; \frac{q}{1-q}P_q^{q-1}(G') - \sum_{i=1}^{K}\theta_i c_i(G') - \alpha \;=\; 0 \qquad \forall G' \in \mathcal{G}_n\,. \tag{A1}$$

To eliminate α, we multiply (A1) by $P_q(G')$ and sum over all $G' \in \mathcal{G}_n$, giving

$$\alpha \;=\; \frac{q}{1-q}\sum_{G\in\mathcal{G}_n} P_q^q(G) \;-\; \sum_{i=1}^{K}\theta_i\langle c_i\rangle\,. \tag{A2}$$

Substituting back Eq. (A2) into Eq. (A1) and multiplying by $(1-q)/q$ yields

$$P_q^{q-1}(G') \;=\; \sum_{G\in\mathcal{G}_n} P_q^q(G) \;+\; \frac{1-q}{q}\sum_{i=1}^{K}\theta_i(c_i(G') - \langle c_i\rangle)\,, \tag{A3}$$

which can be conveniently rewritten as

$$P_q(G') \;=\; \left(\sum_{G\in\mathcal{G}_n} P_q^q(G)\right)^{1/(q-1)} \left[1 - \frac{q-1}{q}\frac{1}{\sum_{G\in\mathcal{G}_n} P_q^q(G)}\left(\sum_{i=1}^{K}\theta_i(c_i(G') - \langle c_i\rangle)\right)\right]_+^{1/(q-1)}. \tag{A4}$$

The prefactor $\left(\sum_{G\in\mathcal{G}_n} P_q^q(G)\right)^{1/(q-1)}$ can be identified with both the inverse partition function $\mathcal{Z}_q^{-1}(\boldsymbol{\theta})$ and the inverse Uffink entropy $U_q^{-1}[P_q]$ of P_q itself – which is an indication of self-referentiality:

$$\mathcal{Z}_q(\boldsymbol{\theta}) \;\equiv\; \left(\sum_{G\in\mathcal{G}_n} P_q^q(G)\right)^{1/(1-q)} \;=\; U_q[P_q]\,. \tag{A5}$$

Now, exploiting the duality relation

$$\exp_q(-x) \;=\; [1 \;-\; (1-q)\,x]_+^{1/(1-q)} \;=\; [1 \;+\; (q-1)\,x]_+^{-1/(q-1)} \;=\; \left[\exp_{2-q}(x)\right]^{-1}, \quad (q\in\mathbb{R})\,, \tag{A6}$$

mapping the q-exponential of $-x$ to the inverse $(2-q)$-exponential of x, the general q-exponential graph probability $P_q(G)$ can be written as

$$\begin{aligned} P_q(G) \;&=\; \frac{1}{\mathcal{Z}_q(\boldsymbol{\theta})}\left[1 - \frac{q-1}{q}\frac{1}{\mathcal{Z}_q^{1-q}(\boldsymbol{\theta})}\left(\sum_{i=1}^{K}\theta_i(c_i(G) - \langle c_i\rangle)\right)\right]_+^{1/(q-1)} \\ &=\; \frac{1}{\mathcal{Z}_q(\boldsymbol{\theta})}\left[\exp_q\left(\frac{1}{q\mathcal{Z}_q^{1-q}(\boldsymbol{\theta})}\sum_{i=1}^{K}\theta_i(c_i(G) - \langle c_i\rangle)\right)\right]^{-1} \qquad \text{(A7)} \\ &=\; \frac{1}{\mathcal{Z}_q(\boldsymbol{\theta})}\exp_{2-q}\left(-\frac{1}{q\mathcal{Z}_q^{1-q}(\boldsymbol{\theta})}\sum_{i=1}^{K}\theta_i(c_i(G) - \langle c_i\rangle)\right). \qquad \text{(A8)} \end{aligned}$$

To resolve the self-referential dependence on both $\mathcal{Z}_q = U_q[P_q]$ and $\langle c_i\rangle$, we use the q-exponential addition formula (cf. [47])

$$\exp_{2-q}(x+y) \;=\; \exp_{2-q}(x)\exp_{2-q}\left(\frac{y}{1+(q-1)x}\right). \tag{A9}$$

With this we can write

$$\exp_{2-q}\Big(-\frac{1}{q\mathcal{Z}_q^{1-q}}\sum_{i=1}^{K}\theta_i(c_i(G)-\langle c_i\rangle)\Big)$$
$$= \exp_{2-q}\left(\frac{1}{q\mathcal{Z}_q^{1-q}}\sum_{i=1}^{K}\theta_i\langle c_i\rangle\right)\exp_{2-q}\left(-\frac{1}{q\mathcal{Z}_q^{1-q}+(q-1)\sum_{i=1}^{K}\theta_i\langle c_i\rangle}\sum_{i=1}^{K}\theta_i c_i(G)\right). \tag{A10}$$

The first factor on the right-hand side is independent of G and therefore cancels in both numerator and denominator of Eq. (A8). We can thus define reparametrized multipliers

$$\tilde{\theta}_i \equiv \frac{\theta_i}{q\mathcal{Z}_q^{1-q}+(q-1)\sum_{i=1}^{K}\theta_i\langle c_i\rangle}, \tag{A11}$$

and obtain

$$P_q(G) = \frac{1}{\mathcal{Z}_q}\exp_{2-q}\left(-\sum_{i=1}^{K}\tilde{\theta}_i c_i(G)\right). \tag{A12}$$

As shown in Ref. [67], the parameters $\tilde{\theta}_i$ can be identified with Casimir invariants associated with the group of reparametrizations of Lagrange multipliers that preserve form invariance of the distribution (up to an arbitrary shift in c_i's). Dropping the tilde and writing $H(G,\boldsymbol{\theta})=\sum_{i=1}^{K}\theta_i c_i(G)$, Eq. (A12) takes the compact form (43).

Appendix B: Large N behavior - detailed computations of the q-exponential Erdős–Rényi model

We derive the two results stated in Sec. IV B: the variational formula for the free entropy density (55) and the large-deviation form of the link-density probability.

Step 1: Bounds on the probability

Since all $\binom{N}{k}$ adjacency matrices with exactly k links are equally weighted by $\exp_{2-q}(-\theta k)$, the probability of observing link density $c=k/N$ is

$$P\left(\frac{L(G)}{N}=c\right) = \binom{N}{Nc}\frac{\exp_{2-q}(-\theta Nc)}{\mathcal{Z}_{q,N}(\theta)}. \tag{B1}$$

We can use standard information-theoretic bounds [40]

$$\frac{1}{N+1}e^{NH_b(c)} \le \binom{N}{Nc} \le e^{NH_b(c)}, \tag{B2}$$

where $H_b(c)=-c\ln c-(1-c)\ln(1-c)$ is the binary entropy. Then let us rewrite the q-exponential as a standard exponential via

$$e^{NH_b(c)}\exp_{2-q}(-\theta Nc) \equiv e^{N\phi_q(c,\theta,N)}, \tag{B3}$$

with the *effective free entropy density* [Eq. (56) of the main text]

$$\phi_q(c,\theta,N) = H_b(c) - \frac{1}{N(1-q)}\ln[1 + (1-q)\theta Nc]. \tag{B4}$$

Using this rewritten form, the bounds become

$$\frac{e^{N\phi_q(c,\theta,N)}}{(N+1)\,\mathcal{Z}_{q,N}(\theta)} \le P\left(\frac{L(G)}{N}=c\right) \le \frac{e^{N\phi_q(c,\theta,N)}}{\mathcal{Z}_{q,N}(\theta)}. \tag{B5}$$

Step 2: Variational formula for Φ_q

Let $c_{\max} \equiv \arg\max_c \phi_q(c,\theta,N)$. We examine each bound in (B5) separately.

- **Lower bound.** Since $P\left(\frac{L(G)}{N}=c\right) \le 1$, the lower bound in (B5) gives $\frac{1}{N+1}e^{N\phi_q(c,\theta,N)} \le \mathcal{Z}_{q,N}(\theta)$. This holds in particular for $c=c_{\max}$; taking logarithms and dividing by N,

$$-\frac{\ln(N+1)}{N} \;+\; \phi_q(c_{\max},\theta,N) \;\le\; \frac{\ln \mathcal{Z}_{q,N}(\theta)}{N}\,. \tag{B6}$$

- **Upper bound.** Summing the upper bound of (B5) over all $N+1$ possible values of c (one for each $k \in \{0,\dots,N\}$) and using $\sum_c P(\cdot)=1$,

$$1 \;\le\; \sum_c \frac{e^{N\phi_q(c,\theta,N)}}{\mathcal{Z}_{q,N}(\theta)} \;\le\; (N+1)\frac{e^{N\phi_q(c_{\max},\theta,N)}}{\mathcal{Z}_{q,N}(\theta)}\,. \tag{B7}$$

  Taking logarithms and dividing by N,

$$\frac{\ln \mathcal{Z}_{q,N}(\theta)}{N} \;\le\; \frac{\ln(N+1)}{N} \;+\; \phi_q(c_{\max},\theta,N)\,. \tag{B8}$$

Since $\ln(N+1)/N \to 0$, squeezing (B6) and (B8) yields[6]

$$\lim_{N\to\infty} \frac{\ln \mathcal{Z}_{q,N}(\theta)}{N} \;=\; \lim_{N\to\infty} \max_{c\in[0,1]} \phi_q(c,\theta,N)\,, \tag{B9}$$

which is Eq. (55) of the main text.

Step 3: Large-deviation form of P

Combining the probability bounds (B5) with the partition function bounds (B6)–(B8) yields

$$\begin{aligned} -\frac{2\ln(N+1)}{N} \;+\; \phi_q(c,\theta,N) \;-\; \phi_q(c_{\max},\theta,N) \;&\le\; \frac{1}{N}\ln P\left(\frac{L(G)}{N}=c\right) \\ &\le\; \phi_q(c,\theta,N) \;-\; \phi_q(c_{\max},\theta,N) \;+\; \frac{\ln(N+1)}{N}\,. \end{aligned} \tag{B10}$$

For fixed θ and q, the second term in (B4) satisfies $\frac{1}{N}\ln(1+(1-q)\theta N c) \to 0$ as $N\to\infty$, so $\phi_q(c,\theta,N) \to H_b(c)$ and $c_{\max} \to 1/2$. Taking $N\to\infty$ in (B10),

$$\lim_{N\to\infty} \frac{1}{N}\ln P\left(\frac{L(G)}{N}=c\right) \;=\; H_b(c) \;-\; H_b(1/2)\,, \tag{B11}$$

or equivalently

$$P\left(\frac{L(G)}{N}=c\right) \;\asymp\; e^{-N[H_b(1/2)-H_b(c)]}\,, \tag{B12}$$

where $a_N \asymp b_N$ means that $\frac{1}{N}(\ln a_N - \ln b_N) \to 0$. This establishes Eq. (61) of the main text, with rate function $I(c) = H_b(1/2) - H_b(c) \ge 0$. Relation (B12) represents the asymptotic equipartition property [40, 77] associated with the Erdős–Rényi model.

---

[6] Strictly, the discrete maximum over $c \in \{0, 1/N, \dots, 1\}$ differs from the continuous maximum over $[0,1]$ by $\mathcal{O}(1/N)$; see Ref. [76].

Appendix C: The average number of links

The average number of links is computed as follows

$$\langle L \rangle \;=\; \frac{\sum_{k=0}^{N} k\binom{N}{k} \exp_{2-q}(-\theta k)}{\mathcal{Z}_q^N(\theta)} \,. \tag{C1}$$

Next, we can use the identity $k\binom{N}{k} = N\binom{N-1}{k-1}$, and the fact that we can write the sum from $k=1$ as the term with $k=0$ does not contribute. The numerator then is

$$\begin{aligned}
\sum_{k=0}^{N} k\binom{N}{k} \exp_{2-q}(-\theta k) \;&=\; N\sum_{k=1}^{N}\binom{N-1}{k-1}(1-(q-1)\theta k)^{1/(q-1)} \\
&=\; N\sum_{k=0}^{N-1}\binom{N-1}{k}(1-(q-1)\theta(k+1))^{1/(q-1)} \\
&=\; N\sum_{k=0}^{N-1}\binom{N-1}{k}(1-(q-1)\theta-(q-1)\theta k)^{1/(q-1)} \\
&=\; N(1-(q-1)\theta)^{1/(q-1)}\sum_{k=0}^{N-1}\binom{N-1}{k}\left(1-(q-1)\frac{\theta}{1-(q-1)\theta}k\right)^{1/(q-1)} \\
&=\; N\exp_{2-q}(-\theta)\,\mathcal{Z}_q^{N-1}\left(\frac{\theta}{1-(q-1)\theta}\right).
\end{aligned} \tag{C2}$$

The average number of links can be then written as

$$\langle L \rangle \;=\; N\exp_{2-q}(-\theta)\,\frac{\mathcal{Z}_q^{N-1}\left(\frac{\theta}{1-(q-1)\theta}\right)}{\mathcal{Z}_q^N(\theta)} \,. \tag{C3}$$

Appendix D: Direct computation of $\langle a_{ij} \rangle$

$$\begin{aligned}
\langle a_{rs} \rangle \;&=\; \frac{1}{\mathcal{Z}_{q,N}(\theta)}\sum_{\{a_{ij}\}} a_{rs}\exp_{2-q}\Big(-\theta\sum_{i=1}^{n}\sum_{j<i}a_{ij}\Big) \;=\; \frac{1}{\mathcal{Z}_{q,N}(\theta)}\sum_{a_{rs}=0,1} a_{rs}\sum_{\substack{\{a_{ij}\}_{i<j}\\(i,j)\neq(r,s)}}\exp_{2-q}\Big(-\theta\sum_{i=1}^{n}\sum_{j<i}a_{ij}\Big) \\
&=\; \frac{1}{\mathcal{Z}_{q,N}(\theta)}\sum_{\substack{\{a_{ij}\}_{i<j}\\(i,j)\neq(r,s)}}\exp_{2-q}\Big(-\theta-\theta\sum_{\substack{i<j\\(i,j)\neq(r,s)}}a_{ij}\Big) \\
&=\; \frac{1}{\mathcal{Z}_{q,N}(\theta)}\exp_{2-q}(-\theta)\sum_{\substack{\{a_{ij}\}_{i<j}\\(i,j)\neq(r,s)}}\exp_{2-q}\Big(-\frac{\theta}{1-(q-1)\theta}\sum_{\substack{i<j\\(i,j)\neq(r,s)}}a_{ij}\Big) \\
&=\; \frac{1}{\mathcal{Z}_{q,N}(\theta)}\exp_{2-q}(-\theta)\mathcal{Z}_{q,N-1}\left(\frac{\theta}{1-(q-1)\theta}\right),
\end{aligned} \tag{D1}$$

where on the third line, we utilized formula (A9), and on the last line we used the definition of $\mathcal{Z}_{q,N}(x)$ [see Eq. (50)].

Appendix E: Computation of higher moments $\langle a_{i_1j_1}\dots a_{i_kj_k} \rangle$

Similarly to how we computed $\langle a_{ij} \rangle$ in (D1), we can compute higher moments. First, let us realize, that $a_{ij} = a_{ij}^2 = a_{ij}^n$ for any $n \in \mathbb{N}$, since we are dealing with binary variables. Therefore, it suffices to consider distinct links.

Let us have k distinct links $i_1 < j_1, \dots, i_k < j_k$. Then we can write

$$\begin{aligned}\langle a_{i_1 j_1} \dots a_{i_k j_k} \rangle &= \frac{1}{\mathcal{Z}_{q,N}(\theta)} \sum_{\{a_{ij}\}} a_{i_1 j_1} \dots a_{i_k j_k} \exp_{2-q}\left(-\theta \sum_{i<j} a_{ij}\right) \\ &= \sum_{a_{i_1 j_1}=0,1} \cdots \sum_{a_{i_k j_k}=0,1} a_{i_1 j_1} \dots a_{i_k j_k} \sum_{\substack{\{a_{ij}\}_{i<j} \\ (i,j)\notin\{(i_\alpha,j_\alpha)\}}} \exp_{2-q}\left(-\theta \sum_{i<j} a_{ij}\right). \end{aligned} \tag{E1}$$

Now, the only summands contributing to the sum are those for which $a_{i_1 j_1} = \dots = a_{i_k j_k} = 1$. Therefore, we can rewrite the sum as

$$\begin{aligned}\langle a_{i_1 j_1} \dots a_{i_k j_k} \rangle &= \frac{1}{\mathcal{Z}_{q,N}(\theta)} \sum_{\substack{\{a_{ij}\}_{i<j} \\ (i,j)\notin\{(i_\alpha,j_\alpha)\}}} \exp_{2-q}\left(-k\theta - \theta \sum_{\substack{i<j \\ (i,j)\notin\{(i_\alpha,j_\alpha)\}}} a_{ij}\right) \\ &= \frac{1}{\mathcal{Z}_{q,N}(\theta)} \exp_{2-q}(-k\theta) \sum_{\substack{\{a_{ij}\}_{i<j} \\ (i,j)\notin\{(i_\alpha,j_\alpha)\}}} \exp_{2-q}\left(-\frac{\theta}{1-k(q-1)\theta} \sum_{\substack{i<j \\ (i,j)\notin\{(i_\alpha,j_\alpha)\}}} a_{ij}\right) \\ &= \frac{1}{\mathcal{Z}_{q,N}(\theta)} \exp_{2-q}(-k\theta)\, \mathcal{Z}_{q,N-k}\left(\frac{\theta}{1-k(q-1)\theta}\right), \end{aligned} \tag{E2}$$

where on the second line we used the identity (A9).

Appendix F: The $q \approx 1$ expansion of $\langle a_{ij} \rangle$ and Cov(a_{ij}, a_{kl})

As discussed, the first and second joint cumulants can be calculated as

$$\left.\frac{\partial \ln \mathcal{Z}_{q,N}}{\partial h_{kl}}\right|_{\{h_{ij}\}=0} = \langle a_{kl} \rangle\,, \tag{F1}$$

$$\begin{aligned}\left.\frac{\partial^2 \ln \mathcal{Z}_{q,N}}{\partial h_{kl} \partial h_{mn}}\right|_{\{h_{ij}\}=0} &= \langle a_{kl} a_{mn} \rangle - \langle a_{kl} \rangle \langle a_{mn} \rangle \\ &= \mathrm{Cov}(a_{kl}, a_{mn})\,, \end{aligned} \tag{F2}$$

where

$$\ln \mathcal{Z}_{q,N}(\theta, \{h_{ij}\}) = \ln \mathcal{Z}_{1,N}(\theta, \{h_{ij}\}) - \frac{1}{2}(q-1)\theta^2 \langle L^2 \rangle_{\mathrm{Sh}} + \mathcal{O}\big((q-1)^2\big)\,. \tag{F3}$$

Since $\mathcal{Z}_{1,N}(\theta, \{h_{ij}\})$ is the Shannonian partition function, we know that

$$\left.\frac{\partial \ln \mathcal{Z}_{1,N}}{\partial h_{kl}}\right|_{\{h_{ij}\}=0} = \langle a_{kl} \rangle_{\mathrm{Sh}}\,, \qquad \left.\frac{\partial^2 \ln \mathcal{Z}_{1,N}}{\partial h_{kl} \partial h_{mn}}\right|_{\{h_{ij}\}=0} = 0\,. \tag{F4}$$

and also

$$\langle a_{ij} \rangle_{\mathrm{Sh}} = \frac{1}{1+e^{\theta-h_{ij}}}\,, \qquad \frac{\partial \langle a_{ij} \rangle_{\mathrm{Sh}}}{\partial h_{ij}} = \frac{e^{\theta-h_{ij}}}{(1+e^{\theta-h_{ij}})^2} = \langle a_{ij} \rangle_{\mathrm{Sh}} \left[1 - \langle a_{ij} \rangle_{\mathrm{Sh}}\right]\,. \tag{F5}$$

We can expand $\langle L^2 \rangle_{\mathrm{Sh}}$ as

$$\begin{aligned}\langle L^2 \rangle_{\mathrm{Sh}} &= \sum_{i<j} \langle a_{ij}^2 \rangle_{\mathrm{Sh}} + \sum_{i<j} \sum_{\substack{k<l \\ (k,l)\neq(i,j)}} \langle a_{ij} a_{kl} \rangle_{\mathrm{Sh}} \\ &= \sum_{i<j} \langle a_{ij} \rangle_{\mathrm{Sh}} + \sum_{i<j} \sum_{\substack{k<l \\ (k,l)\neq(i,j)}} \langle a_{ij} \rangle_{\mathrm{Sh}} \langle a_{kl} \rangle_{\mathrm{Sh}}\,. \end{aligned} \tag{F6}$$

The first derivative with respect to h_{mn}, evaluated at $\{h_{ij}\} = 0$, is

$$\left.\frac{\partial\langle L^2\rangle_{\rm Sh}}{\partial h_{mn}}\right|_{\{h_{ij}\}=0} = \left.\left[\frac{\partial\langle a_{mn}\rangle_{\rm Sh}}{\partial h_{mn}} + 2\sum_{\substack{k<l\\ k\neq m,\, l\neq n}}\langle a_{kl}\rangle_{\rm Sh}\frac{\partial\langle a_{mn}\rangle_{\rm Sh}}{\partial h_{mn}}\right]\right|_{\{h_{ij}\}=0}$$

$$= \frac{e^\theta}{(1+e^\theta)^2} + 2(N-1)\frac{1}{1+e^\theta}\frac{e^\theta}{(1+e^\theta)^2} = \frac{e^\theta}{(1+e^\theta)^3}\left[2N - 1 + e^\theta\right]. \tag{F7}$$

Combining Eqs. (F3), (F4), and (F5), we obtain the final expression in Eq. (91).

For the covariance between distinct links $(m,n) \neq (p,q)$, we compute

$$\left.\frac{\partial^2\langle L^2\rangle_{\rm Sh}}{\partial h_{mn}\partial h_{pq}}\right|_{\{h_{ij}\}=0} = 2\left.\frac{\partial\langle a_{mn}\rangle_{\rm Sh}}{\partial h_{pq}}\right|_{\{h_{ij}\}=0}\left.\frac{\partial\langle a_{pq}\rangle_{\rm Sh}}{\partial h_{mn}}\right|_{\{h_{ij}\}=0} = 2\frac{e^{2\theta}}{(1+e^\theta)^4}. \tag{F8}$$

This implies

$$\mathrm{Cov}(a_{mn}, a_{pq}) = (1-q)\frac{\theta^2 e^{2\theta}}{(1+e^\theta)^4}. \tag{F9}$$

Appendix G: The phase transition condition

To gain further insight into the phase transition described in Sec. IV C, we examine the saddle-point equation (71) in more detail. By rewriting the right-hand side in the form

$$R(c) = \left[1 + \exp\left(\frac{1}{\frac{1}{\theta} + rc}\right)\right]^{-1}, \tag{G1}$$

we note that $R(c)$ is bounded between 0 and 1 and is discontinuous at $c = -1/(r\theta)$, which is negative since both r and θ are positive. It is straightforward to see that on the interval $(-1/(r\theta), +\infty)$, $R(c)$ is a differentiable, monotonically increasing function of c that approaches $1/2$ as $c \to +\infty$. Consequently, the only way for $R(c)$ to intersect the left-hand side $L(c) = c$ more than once is if the derivative of $R(c)$ with respect to c exceeds 1 (the derivative of $L(c)$) over some interval. If this condition is met, there exists a value of θ for which $R(c)$ has multiple intersections with $L(c)$. An illustration is provided in Fig. 11.

To assess this condition more quantitatively, we note that in Eq. (G1), θ acts as a shift along the c-axis and thus does not alter the shape of $R(c)$. Consequently, the above condition is determined solely by the value of r, and it is sufficient to examine the criterion

$$\frac{\partial}{\partial c}\left[1 + \exp\left(\frac{1}{rc}\right)\right]^{-1} > 1. \tag{G2}$$

Let us denote $f(c,r) = \left[1+\exp\left(\frac{1}{rc}\right)\right]^{-1}$ and recognize that it can be written using a *logistic function* $\sigma(x) = \frac{1}{1+e^{-x}}$ as $f(c,r) = \sigma\left(-\frac{1}{rc}\right)$. In passing we note that the logistic function satisfies the following useful identities:

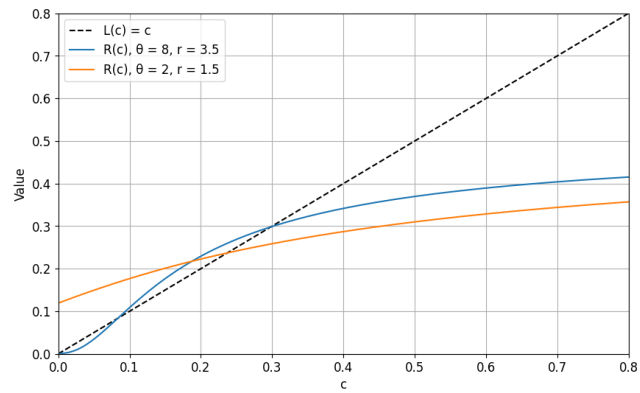

FIG. 11: The left-hand side $L(c) = c$ and the right-hand side $R(c)$ of the saddle-point equation (71) for two distinct combinations of r and θ. To have more than one intersection (solutions to the saddle-point equation), the necessary condition is that the derivative of $R(c)$ must be larger than 1 on some interval.

$$\sigma(-x) = 1 - \sigma(x), \tag{G3}$$

$$\frac{d}{dx}\sigma(x) = \sigma(x)[1-\sigma(x)], \tag{G4}$$

$$2\sigma(x) = 1 + \tanh\left(\frac{x}{2}\right), \tag{G5}$$

$$\tanh\left(\frac{x}{2}\right) = \sigma(x) - \sigma(-x). \tag{G6}$$

Using (G3) and (G4), we rewrite the condition (G2) as

$$\frac{\partial}{\partial c} f(c,r) \;=\; \frac{\partial}{\partial c}\sigma\left(-\frac{1}{rc}\right)$$
$$= \;\sigma\left(-\frac{1}{rc}\right)\sigma\left(\frac{1}{rc}\right)\frac{1}{rc^2} \;>\; 1\,. \qquad \text{(G7)}$$

We recognize that this expression is odd in c, and therefore it is enough to check the condition for $c > 0$. Since we consider $r > 0$, we have $\frac{\partial}{\partial c} f(c,r) > 0$ and it is easy to check that

$$\lim_{c\to 0^+}\frac{\partial}{\partial c} f(c,r) \;=\; \lim_{c\to +\infty}\frac{\partial}{\partial c} f(c,r) \;=\; 0\,. \qquad \text{(G8)}$$

Since $\frac{\partial}{\partial c} f(c,r)$ is differentiable (and hence also continuous), it suffices to check the condition at its maximum.

The condition for the maximum of $\frac{\partial}{\partial c} f(c,r)$ is given by

$$\frac{\partial^2}{\partial c^2} f(c,r) \;=\; 0\,. \qquad \text{(G9)}$$

By evaluating the derivative, we obtain

$$\sigma\left(\frac{1}{rc}\right) \;-\; \sigma\left(-\frac{1}{rc}\right) \;-\; 2rc \;=\; 0\,, \qquad \text{(G10)}$$

and with the help of the identity (G6)

$$\tanh\left(\frac{1}{2rc}\right) \;=\; 2rc\,. \qquad \text{(G11)}$$

The maximum of $\frac{\partial}{\partial c} f(c,r)$ is therefore attained at the value of c satisfying Eq. (G11). The condition (G7) can then be rewritten as

$$r \;>\; \frac{(2rc)^2}{4\sigma\left(-\frac{1}{rc}\right)\sigma\left(\frac{1}{rc}\right)}\,, \qquad \text{(G12)}$$

which can be further rewritten using the identity (G5) and the fact that $\tanh(x)$ is an odd function as

$$r \;>\; \frac{(2rc)^2}{1-\tanh^2\left(\frac{1}{2rc}\right)}\,. \qquad \text{(G13)}$$

Combining Eqs. (G11) and (G13), we see that it suffices to identify a solution x^* of $\tanh(1/x^*) = x^*$. Condition (G13) then implies $r > \frac{(x^*)^2}{1-(x^*)^2}$, which completes the proof.

Appendix H: The critical point

The second derivative of $\phi_r(c,\theta)$ is given by

$$\frac{\partial^2}{\partial c^2}\phi_r(c,\theta) \;=\; \frac{1}{c(c-1)} \;+\; \frac{r}{\frac{1}{\theta}+rc}\,. \qquad \text{(H1)}$$

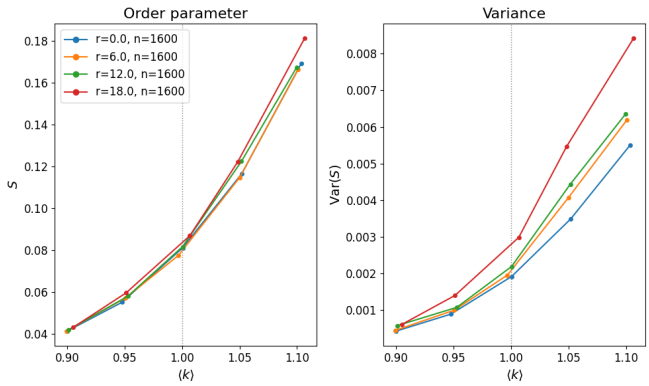

FIG. 12: Order parameter S (left) and its variance $\mathrm{Var}(S)$ (right) as a function of target average degree k^* near the percolation threshold, for $n = 1600$ nodes and $r \in \{0, 6, 12, 18\}$. The mean size of the giant component is insensitive to r, while the variance grows substantially with increasing r, indicating enhanced critical fluctuations.

Setting to zero and rearranging, we get

$$1 \;+\; cr\theta(2-\theta) \;-\; c^2 r\theta^2(r+1) \;=\; 0\,. \qquad \text{(H2)}$$

We want to find the smallest θ, for which this equation has a solution. That is equivalent to finding such θ, for which the discriminant of the quadratic equation is zero. This yields

$$r^2\theta^2(2-\theta)^2 \;-\; 4r\theta^2(r+1) \;=\; 0\,. \qquad \text{(H3)}$$

Assuming $\theta \neq 0$ and using the r value of the critical point r_{cp}, we find

$$\theta_{cp} \;=\; 2 \;+\; \sqrt{\frac{4(r_{cp}+1)}{r_{cp}}}\,. \qquad \text{(H4)}$$

Appendix I: q-ER model – Percolation

For the percolation study, we generated ensembles of 1000 graphs with $n = 1600$ nodes for target average degrees $k^* \in \{0.9, 0.95, 1.0, 1.05, 1.1\}$ (spanning the percolation threshold at $k^* = 1$) and $r \in \{0, 6, 12, 18\}$. We tracked the order parameter S (fraction of nodes in the largest connected component) and its variance $\mathrm{Var}(S)$. As shown in Fig. 12, the mean value of S is virtually identical across all values of r, confirming that the large-N behavior of the giant component is the same as in the Shannon case. However, the variance of S grows monotonically with r: near the percolation threshold, ensembles with larger r exhibit significantly larger fluctuations in the size of the largest connected component. This suggests that while the q-ER model does not shift the percolation threshold, it does amplify the critical fluctuations, consistent with the enhanced link correlations identified analytically in Eq. (92).

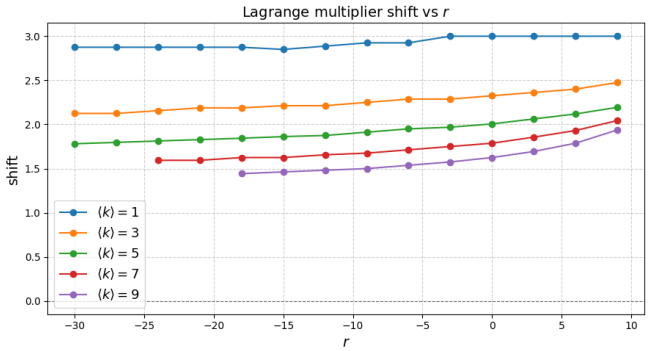

FIG. 13: Lagrange multiplier shift θ_{shift} as a function of r for target average degrees $k^* \in \{1, 3, 5, 7, 9\}$. The shift increases monotonically with r for all target degrees, reflecting that larger $|r|$ requires a stronger compensation to maintain the prescribed density.

Appendix J: q-Exponential Configuration Model – Empirical Results

As described in Sec. V A, we study the q-exponential CM for target average node degrees $k^* \in \{1, 3, 5, 7, 9\}$ and a range of values of $r = (1 - q)N$ from $r = -30$ up to $r = 9$ (for $r > 9$, we often start hitting the irreconstructible region). For each combination of k^* and r, we determine the additive shift θ_{shift} by bisection so that the sampled ensemble reproduces the target density, and then generate 1000 networks with $n = 1000$ nodes using the Metropolis–Hastings algorithm.

1. Fitting the shift parameter

Before comparing structural properties, it is instructive to examine how the required Lagrange multiplier shift θ_{shift} depends on r. As shown in Fig. 13, the shift increases monotonically with r for all target average degrees $k^* \in \{1, 3, 5, 7, 9\}$: moving from negative to positive r (i.e. from $q > 1$ toward $q < 1$) requires a progressively larger additive shift to maintain the target density. Networks with lower target density (note that higher k^* corresponds to a denser graph, so curves for lower k^* sit higher) require a larger absolute shift overall.

2. Degree distributions

An important aspect of our experimental setup is that the parameters θ_i are *not* fitted individually for each node and each value of r. Instead, a single set of θ_i values is sampled from the normal distribution (with $\sigma = 2.28$, $\mu = \sigma^2/2$) and then kept fixed across all values of r; only the common additive shift θ_{shift} is adjusted to recover the target average degree. As a consequence, the degree sequence is not enforced exactly: different values of r induce different link probabilities from the same θ_i profile, leading to differences in the resulting degree distributions.

Fig. 14 shows the complementary cumulative degree distributions for all five target average degrees and a representative subset of r values. For low target degrees ($k^* = 1$), the distributions for different values of r remain nearly indistinguishable. As the target degree increases, however, systematic differences between curves for different r values emerge, particularly in the tail of the distribution. This is a direct consequence of the shared-θ_i setup: at higher densities the non-Shannon coupling has a stronger effect on the individual link probabilities, so the resulting degree sequence deviates more noticeably from the Shannonian baseline as r varies.

3. Tunable clustering

Fig. 15 displays the average clustering coefficient $\langle C \rangle$ as a function of r for each target average degree. A clear and consistent trend is visible across all values of k^*: increasing r (i.e. moving from $q > 1$ toward $q < 1$) leads to a monotonic decrease in average clustering. The range of attainable clustering values widens substantially with increasing k^*, corroborating the result of Sec. V A that the parameter r acts as an independent knob for clustering at fixed degree sequence. This tunability is entirely absent in the standard Shannonian CM, where the degree sequence alone determines the clustering.

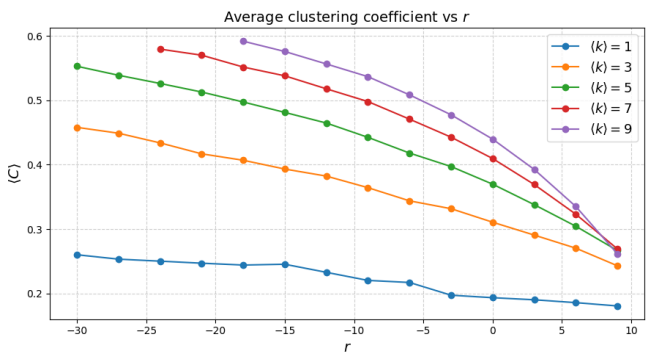

FIG. 15: Average clustering coefficient $\langle C \rangle$ as a function of r for target average degrees $k^* \in \{1, 3, 5, 7, 9\}$. The Shannon ($r = 0$) result is marked for reference. Clustering decreases monotonically with increasing r, and the accessible range grows considerably with k^*, demonstrating that r independently controls the clustering at fixed target density.

4. Degree-resolved network properties

Finally, Fig. 16 shows the average nearest-neighbor degree $\bar{k}_{\text{nn}}(k)$ and the local clustering coefficient $\bar{C}(k)$ as functions of node degree k, for all five target average degrees and several values of r. Both profiles exhibit a systematic shift with varying r: networks with more

negative r (deeper into the $q > 1$ regime) display higher values of both the ANND and the clustering coefficient across all degree classes compared to networks with positive r. The shape of the profiles — both quantities decreasing with k — is consistent with disassortative mixing, but the overall level is controlled by r. This confirms that the q-exponential CM provides a genuinely richer structural repertoire than the Shannonian CM, allowing fine-grained control over higher-order structural properties while respecting the prescribed degree sequence.

Appendix K: Metropolis–Hastings sampling

The Metropolis–Hastings algorithm [78] is a Markov chain Monte Carlo method designed to generate a Markov chain with a prescribed stationary distribution $P(G)$. Starting from a current graph configuration G_t, a candidate graph G' is proposed according to a proposal distribution $Q(G'|G_t)$. The proposed graph is then accepted with probability

$$\alpha \;=\; \min\left(1, \frac{P(G')}{P(G_t)}\frac{Q(G_t|G')}{Q(G'|G_t)}\right), \tag{K1}$$

while G_t is retained with probability $1-\alpha$. A sufficient condition for convergence to the correct stationary distribution is that any graph can be reached from any other in a finite number of proposal steps.

In both the q-ER model and the q-CM, we use a proposal that flips a uniformly at random chosen link (i.e. toggles a non-existent link to present, or an existing link to absent). This proposal is symmetric, $Q(G'|G) = Q(G|G')$, so the acceptance probability reduces to

$$\alpha \;=\; \min\left(1, \frac{P(G')}{P(G_t)}\right). \tag{K2}$$

Crucially, only the ratio of graph probabilities enters, so the (intractable) partition function cancels.

Because each update changes only a single link, successive graphs in the chain are strongly correlated. To reduce autocorrelations we apply *thinning*: after every M proposed steps — where M is of the order of the total number of possible links — we append one graph to the ensemble. We generate ensembles of 1000 graphs for each set of model parameters studied in this work.

Appendix L: Link probability in the q-exponential CM

We derive Eq. (97) of the main text. Starting from the definition and isolating the sum over a_{kl},

$$\begin{aligned}
\langle a_{kl}\rangle \mathcal{Z}_q(\theta_1,\ldots,\theta_N) \;&=\; \sum_{\{a_{ij}\}} a_{kl}\exp_{2-q}\Big[-\sum_{i=1}^{n}\sum_{j<i}(\theta_i+\theta_j)a_{ij}\Big] \\
&=\; \sum_{a_{kl}=0,1} a_{kl} \sum_{\substack{\{a_{ij}\}\\(i,j)\neq(k,l)}} \exp_{2-q}\Big[-(\theta_k+\theta_l)a_{kl} \;-\; \sum_{\substack{i<j\\(i,j)\neq(k,l)}}(\theta_i+\theta_j)a_{ij}\Big] \\
&=\; \sum_{\substack{\{a_{ij}\}\\(i,j)\neq(k,l)}} \exp_{2-q}\Big[-(\theta_k+\theta_l) \;-\; \sum_{\substack{i<j\\(i,j)\neq(k,l)}}(\theta_i+\theta_j)a_{ij}\Big] \\
&=\; \exp_{2-q}(-(\theta_k+\theta_l)) \sum_{\substack{\{a_{ij}\}\\(i,j)\neq(k,l)}} \exp_{2-q}\left[-\sum_{\substack{i<j\\(i,j)\neq(k,l)}} \frac{\theta_i+\theta_j}{1-(q-1)(\theta_k+\theta_l)}a_{ij}\right] \\
&\equiv\; \exp_{2-q}(-(\theta_k+\theta_l))\,\tilde{\mathcal{Z}}_q^{(k,l)}\Big(\tilde{\theta}_1^{(k,l)},\ldots,\tilde{\theta}_N^{(k,l)}\Big),
\end{aligned} \tag{L1}$$

where the third-to-fourth line uses the q-exponential addition formula (A9). Dividing both sides by $\mathcal{Z}_q$ yields Eq. (97).

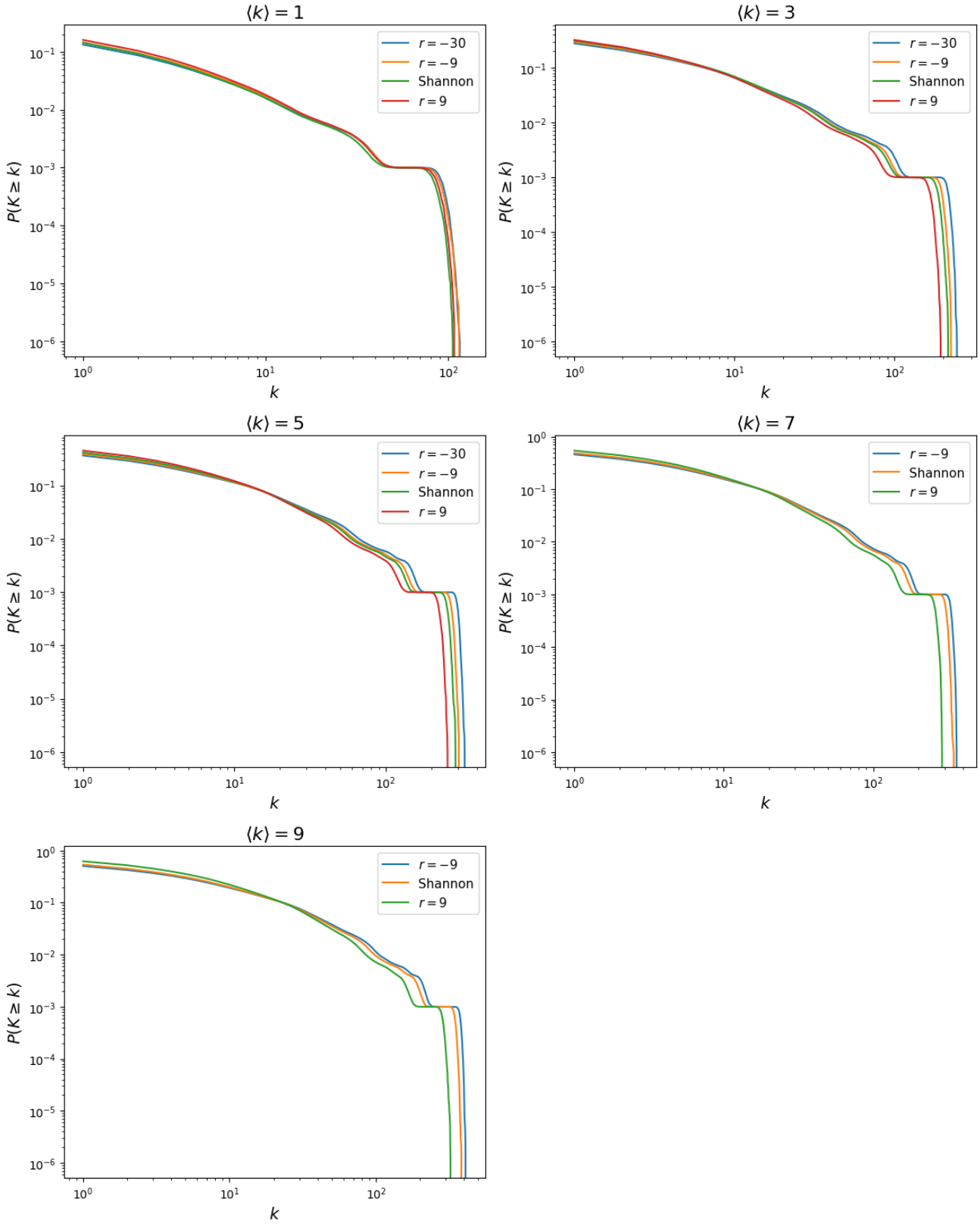

FIG. 14: Complementary cumulative degree distributions for the q-exponential CM at target average degrees $k^* \in \{1, 3, 5, 7, 9\}$ and selected values of $r \in \{-30, -9, \text{Shannon}, 9\}$. Note that the parameters θ_i are sampled once and held fixed across all r; only the common shift θ_{shift} is tuned to match the target average degree. Distributions for different r values are nearly identical at low target degrees but begin to diverge in the tail as the target degree increases, reflecting the growing influence of the non-Shannon coupling on the individual link probabilities.

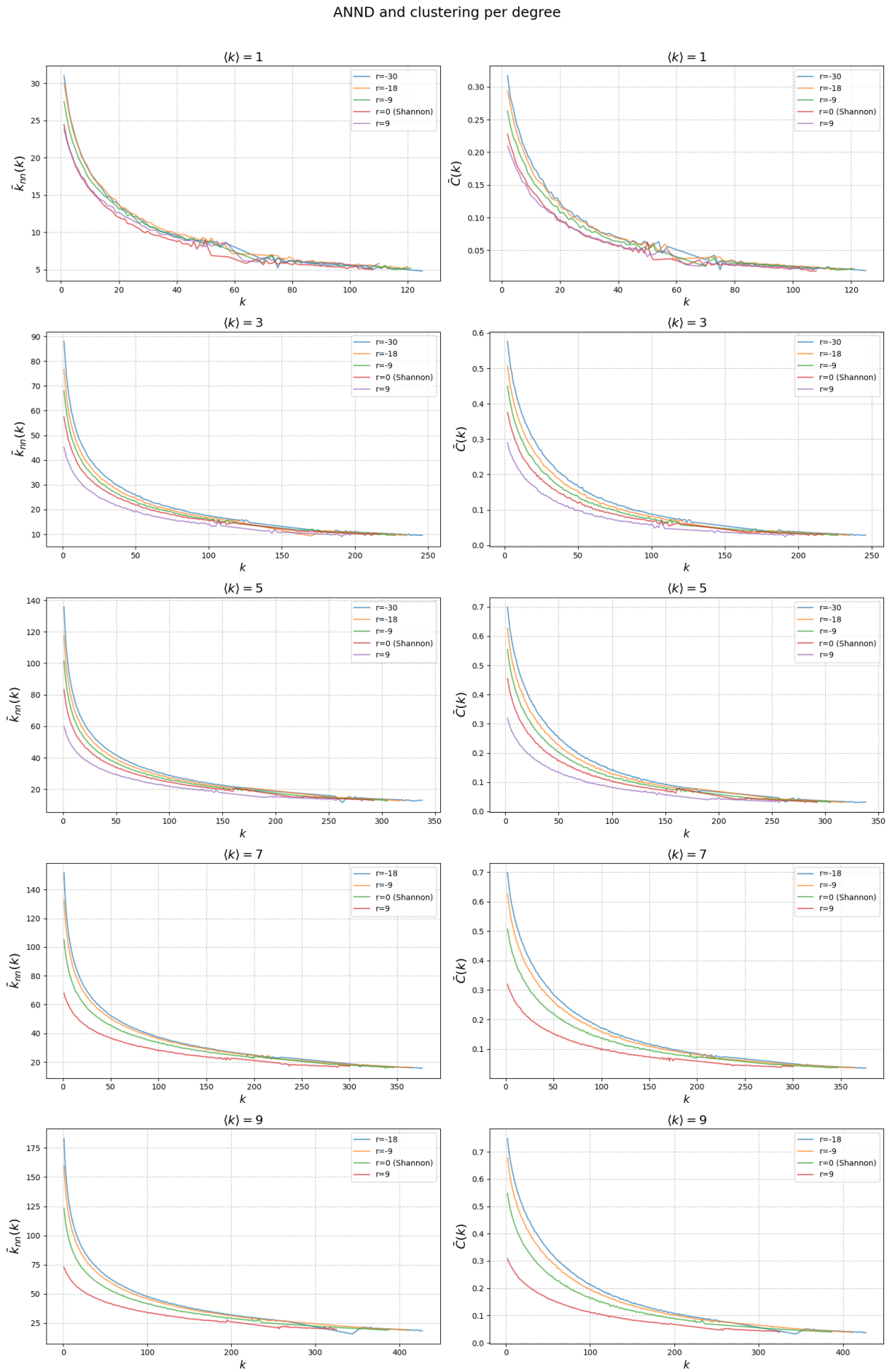

FIG. 16: Average nearest-neighbor degree $\bar{k}_{\mathrm{nn}}(k)$ (left column) and local clustering coefficient $\bar{C}(k)$ (right column) as functions of node degree k, for target average degrees $k^* \in \{1, 3, 5, 7, 9\}$ (rows) and selected values of r. Both quantities decrease with k in all cases, but their overall level shifts systematically with r, demonstrating that the non-Shannon parameter independently controls higher-order structural properties beyond the degree sequence.

[1] P. W. Holland and S. Leinhardt, Journal of the American Statistical Association **76**, 33 (1981), ISSN 1537-274X.
[2] D. Strauss, SIAM review **28**, 513 (1986).
[3] S. Wasserman, *Social network analysis: Methods and applications*, vol. 8 (Cambridge university press, 1994).
[4] C. J. Anderson, S. Wasserman, and B. Crouch, Social networks **21**, 37 (1999).
[5] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, Social networks **29**, 173 (2007).
[6] J. Park and M. E. J. Newman, Physical Review E **70**, 066117 (2004), ISSN 1550–2376.
[7] G. Bianconi, Physical Review E **79**, 036114 (2009).
[8] A. Fronczak, arXiv preprint arXiv:1210.7828 (2012).
[9] T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics* (Springer International Publishing, 2017), ISBN 9783319694382.
[10] A. C. Coolen, A. Annibale, and E. S. Roberts, *Generating random networks and graphs* (Oxford university press, 2017).
[11] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, Nature Reviews Physics **1**, 58 (2019), ISSN 2522–5820.
[12] C. E. Shannon, Bell System Technical Journal **27**, 379 (1948), ISSN 0005-8580.
[13] P. Erdős and A. Rényi, Publicationes Mathematicae Debrecen **6**, 290–297 (1959).
[14] P. Erdős, A. Rényi, et al., Publications of the (1960).
[15] J. Park and M. E. J. Newman, Phys. Rev. E **68**, 026112 (2003).
[16] P. Fronczak, A. Fronczak, and M. Bujok, Physical Review E **88** (2013).
[17] B. Karrer and M. E. J. Newman, Physical Review E **83**, 016107 (2011).
[18] T. P. Peixoto, Advances in network clustering and blockmodeling pp. 289–332 (2019).
[19] C. J. Anderson, S. Wasserman, and K. Faust, Social networks **14**, 137 (1992).
[20] D. Garlaschelli and M. I. Loffredo, Physical Review Letters **93**, 268701 (2004).
[21] T. Squartini, F. Picciolo, F. Ruzzenenti, and D. Garlaschelli, Scientific reports **3**, 2729 (2013).
[22] A. Gallo, F. Saracco, R. Lambiotte, D. Garlaschelli, and T. Squartini, Physical Review E **111**, 024312 (2025).
[23] M. E. J. Newman, SIAM Review **45**, 167 (2003).
[24] P. Fronczak, A. Fronczak, and J. A. Hołyst, The European Physical Journal B **59**, 133 (2007).
[25] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[26] P. Holme, Physical Review E—Statistical, Nonlinear, and Soft Matter Physics **72**, 046111 (2005).
[27] M. E. Newman, Proceedings of the national academy of sciences **103**, 8577 (2006).
[28] F. Saracco, M. J. Straka, R. D. Clemente, A. Gabrielli, G. Caldarelli, and T. Squartini, New Journal of Physics **19**, 053022 (2017).
[29] A. Gallo, F. Saracco, and T. Squartini, npj Complexity **2**, 22 (2025).
[30] D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris, Journal of statistical software **24**, 1 (2008).
[31] T. Squartini and D. Garlaschelli, New Journal of Physics **13**, 083001 (2011), ISSN 1367-2630.
[32] T. Squartini, G. Caldarelli, G. Cimini, A. Gabrielli, and D. Garlaschelli, Physics Reports **757**, 1 (2018), ISSN 0370-1573.
[33] G. Cimini, R. Mastrandrea, and T. Squartini, *Reconstructing networks* (Cambridge University Press, 2021).
[34] R. W. Krause, M. Huisman, C. Steglich, and T. A. Snijders, in *2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)* (IEEE, 2018), pp. 159–163.
[35] E. T. Jaynes, Physical Review **106**, 620 (1957), ISSN 0031-899X.
[36] E. T. Jaynes, Physical Review **108**, 171 (1957), ISSN 1536-6065.
[37] J. E. Shore and R. W. Johnson, IEEE Transactions on Information Theory **26**, 26 (1980).
[38] J. E. Shore and R. W. Johnson, IEEE Trans. Inf. Theory **27**, 472 (1981).
[39] A. Khinchin, *Mathematical Foundations of Information Theory* (Dover Publications, New York,, 1957).
[40] T. M. Cover and J. A. Thomas, *Elements of information theory* (Wiley-Interscience, Hoboken, N.J, 2006), second edition ed., ISBN 9780471748823, includes bibliographical references (pages 689-721) and index.
[41] J. W. Gibbs, *Elementary principles in statistical mechanics* (Courier Corporation, 2014).
[42] L. Boltzmann, Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften in Wien **66**, 275 (1872).
[43] J. Uffink, Stud. Hist. Phil. Mod. Phys. **26**, 223 (1995).
[44] P. Jizba and J. Korbel, Physical Revivew Letters **122**, 120601 (2019).
[45] P. Jizba and J. Korbel, Physical Review E **101**, 042126 (2020).
[46] S. Thurner, R. Hanel, and P. Klimekl, *Introduction to the Theory of Complex Systems* (Oxford University Press, 2018), ISBN 9780198821939.
[47] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics*, SpringerLink (Springer New York, New York, NY, 2009), ISBN 9780387853598.
[48] C. Tsallis, Journal of Statistical Physics **52**, 479 (1988), ISSN 1572-9613.
[49] A. Rényi, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (University of California Press, 1961), pp. 547–561.
[50] A. Rényi, *Selected Papers of Alfred Rényi, Vol. 2* (Akadémia Kiado, 1976).
[51] A. N. Kolmogorov and G. Castelnuovo, *Sur la notion de la moyenne* (G. Bardi, tip. della R. Accad. dei Lincei, 1930).
[52] M. Nagumo, in *Japanese journal of mathematics: transactions and abstracts* (The Mathematical Society of Japan, 1930), vol. 7, pp. 71–79.
[53] A. Somazzi and D. Garlaschelli, Physical Review Research **7**, 033087 (2025).
[54] P. Jizba and J. Korbel, Physical Review E **100**, 026101 (2019).
[55] P. A. Morales and F. E. Rosas, Physical Review Research **3**, 033216 (2021).
[56] F. Battiston and G. Petri, *Higher-order systems*

(Springer, 2022).
[57] C. Bick, E. Gross, H. A. Harrington, and M. T. Schaub, SIAM review **65**, 686 (2023).
[58] G. Bianconi, *Higher-order networks* (Cambridge University Press, 2021).
[59] S. Boccaletti, P. De Lellis, C. I. del Genio, K. Alfaro-Bittner, R. Criado, S. Jalan, and M. Romance, Physics Reports **1018**, 1 (2023).
[60] T. P. Peixoto, L. Peel, T. Gross, and M. De Domenico, arXiv preprint arXiv:2602.16937 (2026).
[61] T. A. Snijders et al., Journal of Social Structure **3**, 1 (2002).
[62] N. Vallarano, M. Bruno, E. Marchese, G. Trapani, F. Saracco, G. Cimini, M. Zanon, and T. Squartini, Scientific Reports **11**, 15227 (2021).
[63] D. Garlaschelli and M. I. Loffredo, Physical Review E—Statistical, Nonlinear, and Soft Matter Physics **78**, 015101 (2008).
[64] E. M. F. Curado and C. Tsallis, Journal of Physics A: Mathematical and General **24**, L69 (1991).
[65] C. Tsallis, R. D. Mendes, and A. R. Plastino, Physica A: Statistical Mechanics and its Applications **261**, 534 (1998), ISSN 0378-4371.
[66] G. L. Ferri, S. S. Martínez, and A. R. Plastino, Journal of Statistical Mechanics: Theory and Experiment **2005**, P04009 (2005), ISSN 1742-5468.
[67] P. Jizba, J. Korbel, and V. Zatloukal, Physical Review E **95**, 022103 (2017).
[68] J. Park and M. E. Newman, Physical Review E—Statistical, Nonlinear, and Soft Matter Physics **70**, 066146 (2004).
[69] A. Annibale and O. T. Courtney, Journal of Physics A: Mathematical and Theoretical **48**, 365001 (2015).
[70] J. Park and M. E. Newman, Physical Review E—Statistical, Nonlinear, and Soft Matter Physics **72**, 026136 (2005).
[71] H. Touchette, Physical Reports **478**, 1 (2009).
[72] R. Hanel and S. Thurner, EPL (Europhysics Letters) **93**, 20006 (2011), ISSN 1286-4854.
[73] A. Gabrielli, V. Macchiati, and D. Garlaschelli, *Critical Density for Network Reconstruction* (Springer Nature Switzerland, 2024), pp. 223–249, ISBN 9783031552489.
[74] M.-Á. Serrano and M. Boguná, Physical Review E—Statistical, Nonlinear, and Soft Matter Physics **72**, 036133 (2005).
[75] J. Korbel, R. Hanel, and T. Stefan, New Journal of Physics **20**, 093007 (2018).
[76] A. Dembo and A. Montanari, Brazilian Journal of Probability and Statistics **24** (2010), ISSN 0103-0752.
[77] B. Corominas-Murtra, R. Hanel, and P. Jizba, The European Physical Journal B **97**, 129 (2024).
[78] C. P. Robert and G. Casella, *The Metropolis—Hastings Algorithm* (Springer New York, 1999), pp. 231–283, ISBN 9781475730715.