

Entropy of labeled versus unlabeled networks

Jeremy Paton^{1,2}, Harrison Hartle^{1,2}, Huck Stepanyants^{1,2}, Pim van der Hoorn³, and Dmitri Krioukov^{1,2,4,5}

¹Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA

²Network Science Institute, Northeastern University, Boston, Massachusetts 02115, USA

³Department of Mathematics and Computer Science, Eindhoven University of Technology, 5600 MB Eindhoven, Netherlands

⁴Department of Mathematics, Northeastern University, Boston, Massachusetts 02115, USA

⁵Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115, USA



(Received 20 April 2022; accepted 24 October 2022; published 17 November 2022)

The structure of a network is an unlabeled graph, yet graphs in most models of complex networks are labeled by meaningless random integers. Is the associated labeling noise always negligible, or can it overpower the network-structural signal? To address this question, we introduce and consider the sparse unlabeled versions of popular network models and compare their entropy against the original labeled versions. We show that labeled and unlabeled Erdős-Rényi graphs are entropically equivalent, even though their degree distributions are very different. The labeled and unlabeled versions of the configuration model may have different prefactors in their leading entropy terms, although this remains conjectural. Our main results are upper and lower bounds for the entropy of labeled and unlabeled one-dimensional random geometric graphs. We show that their unlabeled entropy is negligible in comparison with the labeled entropy. This means that in sparse networks the entropy of meaningless labeling may dominate the entropy of the network structure. The main implication of this result is that the common practice of using exchangeable models to reason about real-world networks with distinguishable nodes may introduce uncontrolled aberrations into conclusions made about these networks, suggesting a need for a thorough reexamination of the statistical foundations and key results of network science.

DOI: [10.1103/PhysRevE.106.054308](https://doi.org/10.1103/PhysRevE.106.054308)

I. INTRODUCTION

Networks are everywhere, and all of them are labeled. The labels of people in social networks are their names and all their other metadata, such as occupation, interests, place of living, and so on. Similarly, genes in gene regulatory networks, routers in the Internet, or countries in the world trade web all have their unique meaningful names or identifiers. This labeling information is ignored if one is interested in the structure of a real-world network. Here we assume that the *structure* of a network is defined to be an *unlabeled* graph, many visualizations of which can be found in textbooks, papers, or presentations in network science and graph theory.

Yet all of the popular network models used to study the structure of real-world networks are not unlabeled. Nodes in these models are actually labeled. However, there is a drastic difference between node labels in real-world networks and node labels in network models. Since network models are typically abstract mathematical models of random graphs, node labels in them cannot be as meaningful as the names of countries in the world, for instance. Node labels in network models come from arbitrary abstract sets of size n , the network size. Without loss of generality, such label sets can be and usually are set to the set of integers from 1 to n , denoted by $[n] := \{1, 2, \dots, n\}$. Furthermore, since graphs in these models are typically random, so are labels in them. Any node in any Erdős-Rényi graph of size 100 can have label 99, for instance, as opposed to the label *Bhutan* attached to an individual country in the world trade web.

Can the entropy coming from such random labeling of an unlabeled graph, the network structure, be safely ignored, or can it introduce non-negligible aberrations into the system that we have to account for in a nontrivial way? In other words, since the meaningless random labels $[n]$ are nothing but “noise,” assigned to an unlabeled graph uniformly at random out of the $n!$ permutations, then can it be the case that this noise statistically dominates the randomness associated with the network structure, an unlabeled graph?

The principled way to address this question is to compare the leading terms of network entropy in the labeled and unlabeled cases. If the former dominates the latter, then indeed the noise overpowers the signal. The other reason to focus on entropy is that entropy is one of the most important properties of a network model, a central player in the definitions of the unbiased null models of networks [1–13], network ensemble equivalence [5,14–16], network typicality [17,18], and many other fundamental matters [19–30]. Yet our main motivation are null models, which are network models that maximize network entropy under various network-structural constraints.

Here we show that maximizing the entropy of the naked network structure represented by an unlabeled graph, and maximizing the entropy of this structure dressed in random labels represented by a labeled graph, may lead to very different outcomes in sparse networks. The entropy of meaningless random labeling may be a dominating factor, so it gets maximized, instead of the intended maximization of network-structural entropy.

TABLE I. The leading terms of labeled and unlabeled entropies of the sparse microcanonical Erdős-Rényi random graphs (ER), configuration model with scale-free degree sequences (CM), and one-dimensional random geometric graphs (RGG). The scaling of unlabeled CM entropy is conjectural.

| Entropy | ER | CM | RGG |
|-----------|-------------------------------|-----------------------------------|-----------------|
| Labeled | $\approx (\bar{k}/2)n \log n$ | $\approx (\bar{k}/2)n \log n$ | $\sim n \log n$ |
| Unlabeled | $\approx (\bar{k}/2)n \log n$ | $\approx (\bar{k}/2 - 1)n \log n$ | $\lesssim n$ |

In what follows, we first recall in Sec. II the key differences between labeled and unlabeled graphs and explain why in most cases—essentially in all the cases that deal with maximum-entropy null models of real networks—one should be interested in unlabeled graph models, versus their well-known labeled counterparts. In a nutshell, this is because labels in real networks are “glued” to individual nodes, like *Bhutan* to the country, resulting in one network structure, as opposed to labeled network models where this structure is repeated as many times as the number of graph isomorphisms, i.e., the number of label permutations leading to a different but isomorphic labeled graph. As a consequence, labeled models are biased toward more asymmetric graphs, with larger isomorphism classes, compared to the unbiased unlabeled models of the network structure with the same sufficient statistics.

In Sec. III, we then consider the two most basic illustrative examples: (1) the “harmonic oscillator” of network models—the Erdős-Rényi random graphs (ER) and (2) the configuration model (CM). In the ER case, we consider the unlabeled versions of both microcanonical $\mathcal{G}_{n,m}$ and canonical $\mathcal{G}_{n,p}$ labeled ER graphs, denoting these unlabeled ER models by $\mathcal{U}_{n,m}$ and $\mathcal{U}_{n,p}$. Somewhat shockingly, the unlabeled canonical ER graphs $\mathcal{U}_{n,p}$ have neither been considered nor even properly defined before. It is known, however, that even such a basic property as the degree distribution is very different in sparse microcanonical labeled $\mathcal{G}_{n,m}$ versus unlabeled $\mathcal{U}_{n,m}$ ER graphs—by *sparse graphs* we mean graphs with constant average degree $\bar{k} = 2m/n$. Notwithstanding these differences, we show that the leading term of entropy of both $\mathcal{G}_{n,m}$ and $\mathcal{U}_{n,m}$ is surprisingly the same $(\bar{k}/2)n \log n$. The subleading entropy terms are different, however. The leading term of entropy of the unlabeled microcanonical CM with scale-free degree distributions is unknown but is not excluded to be the same as in the labeled case, albeit with a different prefactor $(\bar{k}/2 - 1)n \log n$. The unlabeled canonical CM has never been mentioned before either, so we define it in Sec. III as well.

Our main results are in Sec. IV. They are tight lower and upper bounds for the entropy of unlabeled and labeled one-dimensional random geometric graphs (RGGs). The calculation of RGG entropy is an important longstanding problem that has seen only limited progress as it has been considered intractable [31–34]. We develop a powerful technique, rooted in the labeled-unlabeled delineation, that allows us to show that the leading terms of the entropy of sparse unlabeled and labeled RGGs are different, $\lesssim n$ versus $\sim n \log n$, respectively, Table I. This disconcerting result implies that the entropy of labeled graphs is dominated by the entropy of the

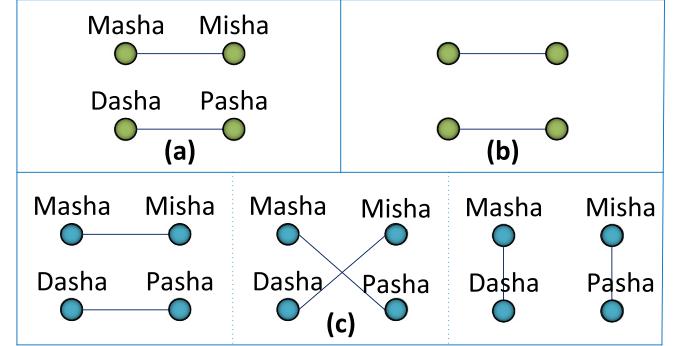


FIG. 1. A network of love: a “real-world” network (a) and its unlabeled (b) and exchangeably labeled (c) versions.

meaningless labeling noise rather than by the entropy of the network structure.

The entropic equivalence is a necessary but not sufficient condition for the ensemble equivalence [5,14–16]. Our RGG result thus says that labeled and unlabeled RGGs are statistically very different. Yet even if the leading entropy terms are the same, as in labeled and unlabeled ER graphs, the ensembles can still be very different, up to the point that their degree distributions can be very different. The main overall conclusions are then that unlabeled network models may behave very differently from their labeled counterparts, so any predictions concerning the network structure based on labeled models may lead to potentially misleading or statistically incorrect outcomes. These and other implications and challenges are discussed in Sec. V.

A. Notations and conventions

In what follows, the symbols “ \ll ,” “ \sim ,” “ \approx ,” and “ \gg ” in $a_n \sim b_n$ mean that $c = \lim_{n \rightarrow \infty} a_n/b_n$ is $c = 0$, $0 < c < \infty$, $c = 1$, and $c = \infty$, respectively. We call networks sparse or dense if their (expected) average degree is $\bar{k} \sim 1$ or $\bar{k} \sim n$. All networks are sparse below, unless mentioned otherwise. The adjunctive *with high probability* is implied where needed.

II. LABELED VS. UNLABELED NETWORKS

Consider a “real-world” love network among the four people in Fig. 1(a). Masha loves Mishka while Dasha loves Pasha. The nodes are labeled by lovers’ names, so the graph is labeled. What permutations of labels are allowed in this network? Clearly, we can only swap Masha with Mishka and/or Dasha with Pasha. These swaps are called graph *automorphisms*: Before and after the swap the network is the same labeled graph. No other permutation of labels is allowed. We cannot swap Masha with Dasha or with Pasha, for instance, because such label permutations lead to different labeled graphs, with different love stories in the real life. Labels in real-world networks are thus “glued” to nodes since nodes are *distinguishable* entities.

It is now critical to recognize that because of this gluing—or more formally, since no non-automorphism label permutations are allowed in real networks—we are essentially dealing with *unlabeled* networks. Indeed, if we are not concerned who exactly loves whom exactly—that is, if we are

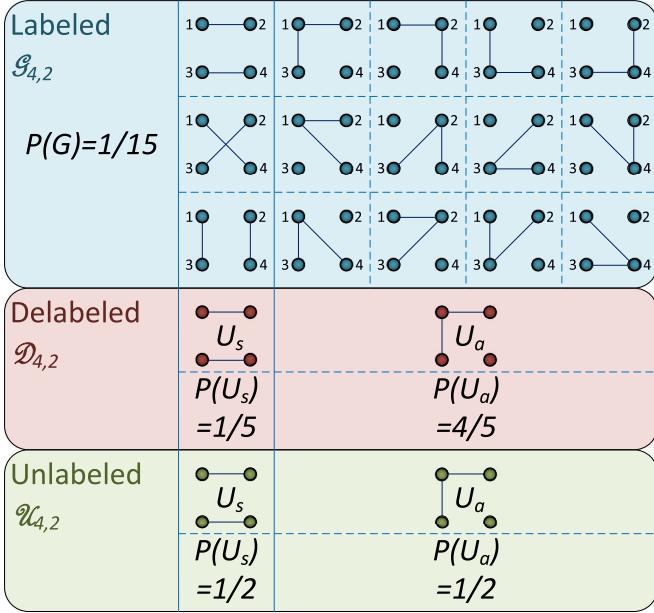


FIG. 2. The labeled, delabeled, and unlabeled microcanonical Erdős-Rényi random graphs of size $n = 4$ with $m = 2$ edges. The probabilities of all labeled graphs G are the same $P(G) = 1/15$ in $\mathcal{G}_{n,m}$, while the probabilities of the symmetric and asymmetric unlabeled graphs U_s and U_a are different in the delabeled and unlabeled models $\mathcal{D}_{n,m}$ and $\mathcal{U}_{n,m}$.

interested only in the network *structure*—then the network structure in the example is a pair of couples, represented by the *unlabeled* graph in Fig. 1(b). If we are interested in who loves whom, then again it is just *one labeled* graph in Fig. 1(a). In either case, we are dealing with just *one* graph, either labeled or unlabeled, and not with the *three* isomorphic graphs labeled in all the three possible ways in Fig. 1(c). (Two labeled graphs are called *isomorphic* if they are the same unlabeled graph.) Only *one* of these three labeled graphs reflects reality; the other two are “noise.”

Unlabeled graphs can but do not have to be considered as isomorphism classes of labeled graphs. In fact, representing unlabeled graphs as isomorphism classes of labeled graphs can be confusing. The easiest way to understand unlabeled graphs is via their enumeration [35], a simple example of which we consider next.

Suppose we are to formulate the statistically correct null model of networks with $n = 4$ nodes and $m = 2$ edges, as in our example with Misha, Masha, Pasha, and Dasha. By *statistically correct* models we mean here the unbiased models that maximize entropy subject to given constraints [1–13]. In our example, these constraints are $n = 4$ and $m = 2$, so the correct entropy-maximizing null models are *defined* by the uniform distributions over the space of all graphs with $n = 4$ nodes and $m = 2$ edges. However, this space is very different for unlabeled versus labeled graphs. There are only two unlabeled graphs with 4 nodes and 2 edges, while there are 15 labeled ones, all shown in Fig. 2. The uniform distribution $P(G) = 1/15$ over the 15 labeled graphs G is the familiar Erdős-Rényi model $\mathcal{G}_{n,m}$ with $n = 4$ and $m = 2$, while the uniform distribution $P(U) = 1/2$ over the two unlabeled

graphs U is its virtually unknown unlabeled counterpart $\mathcal{U}_{n,m}$. Which one, $\mathcal{G}_{n,m}$ or $\mathcal{U}_{n,m}$, are we supposed to work with in applications to real networks?

The answer to this question depends on whether the nodes in a real network are distinguishable or indistinguishable. The most crucial difference between the considered “real” love network and the labeled ER $\mathcal{G}_{4,2}$ is that the nodes in the love network are distinguishable, while they are indistinguishable in $\mathcal{G}_{4,2}$. Indeed, the label set in $\mathcal{G}_{4,2}$ is not {Misha, Masha, Pasha, Dasha} but {1, 2, 3, 4}, and the model is exchangeable. A model of labeled graphs is called *exchangeable* if the probabilities of any two isomorphic graphs in the model are the same. Exchangeability is thus a statistical formalization of the idea that node labels “do not matter” and can be permuted arbitrarily. In other words, *exchangeability is a formalization of statistical indistinguishability*.

Therefore, the very common practice of applications of *exchangeable* models of labeled graphs to real networks with *distinguishable* nodes is statistically questionable. Either *nonexchangeable* models, in which node labels “do matter and are glued to nodes,” must be used for such networks, or—as far as null models of network structure are concerned—the statistically correct null models of such networks must be models of *unlabeled* graphs. In either case, we are dealing not with three graphs in Fig. 1(c) but with one graph, either in Fig. 1(a) or in Fig. 1(b).

The other way around, if nodes in a real-world network are *indistinguishable*, such as atoms in material networks [36], then the statistically correct null models of such networks must be *exchangeable* models of *labeled* graphs. Indeed, different atoms are different atoms, but since they are indistinguishable, any permutation of individual atoms in a particular configuration is equally good, a typical situation in statistical physics [37], perhaps the most vivid illustration of which is the infamous Gibbs paradox [38]. However, such situations in the science of real-world complex networks are rare exclusions rather than a rule, since in a vast majority of real networks, nodes are distinguishable.

But does it really matter which models to use, labeled or unlabeled, as they may be equivalent in some way? The key message of this paper is that it really does matter what models we deal with as the labeled and unlabeled versions of the same model can be very nonequivalent. We already see clear signs of this in our toy example with $\mathcal{G}_{4,2}$ and $\mathcal{U}_{4,2}$ in Fig. 2. The two models are clearly very different in many respects. For instance, the probability of the asymmetric U_a scenario in Fig. 2 with one lucky person loving two others, while the unlucky fourth is left loveless, is 80% in the labeled $\mathcal{G}_{4,2}$, while it is only 50% in the unlabeled $\mathcal{U}_{4,2}$. That is, the two models give different predictions concerning the likelihoods of different “love scenarios in the real life.”

The last observation illustrates the role of the symmetry of a graph in analyzing the statistical differences between labeled and unlabeled graph models. The graph U_a in Fig. 2 is “more asymmetric” than the pair of couples U_s because there are more label permutations on U_a ’s labeled version that lead to different labeled graphs. In graph theory, a graph is called *asymmetric* if any permutation of its labels is not an automorphism, i.e., if any label permutation leads to a different labeled graph. The graph is called *symmetric* otherwise. The

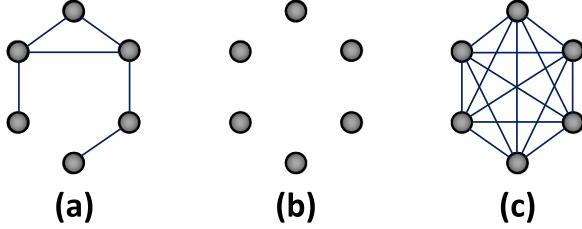


FIG. 3. The graph symmetry extremes: the smallest and sparsest fully asymmetric graph (a) and the fully symmetric empty (b) and complete (c) graphs of the same size.

smallest asymmetric graphs are of size 6, and there are eight such graphs. The single one with the fewest edges (6) is shown in Fig. 3(a). In $\mathcal{U}_{6,6}$, it is just one unlabeled graph, but in $\mathcal{G}_{6,6}$, it corresponds to $6! = 720$ labeled graphs, in stark contrast with the fully symmetric empty or complete graphs, Figs. 3(b) and 3(c), represented by only one graph, either unlabeled or labeled, since any label permutation of an empty or complete graph is an automorphism.

These observations are directly related to the common confusion between *unlabeled* graph models and *delabeled* graph models [39,40]. A delabeled graph model starts with a labeled graph model, generates a labeled graph, and then simply removes the node labels in it. The result is a model of unlabeled graphs, in which more asymmetric graphs attract higher probability masses. In our example above, the delabeled ER model $\mathcal{D}_{n,m}$ with $n = 4$ nodes and $m = 2$ edges is the probability distribution on the two unlabeled graphs in Fig. 2, which assigns the probability of $12/15 = 80\%$ to the freaky asymmetric scenario U_a , and only $3/15 = 20\%$ to the more conventional pair of couples U_s , as opposed to the unlabeled $\mathcal{U}_{n,m}$, which says the two scenarios are equally likely, the probability of each is $1/2 = 50\%$. In other words, if you generate a random $\mathcal{G}_{n,m}$ graph (e.g., by placing m edges randomly among $\binom{n}{2}$ node pairs), and then consider the generated graph as unlabeled, then you have sampled a random unlabeled graph *not* from the unlabeled model $\mathcal{U}_{n,m}$ but from the delabeled model $\mathcal{D}_{n,m}$. This shows that as far as the probability of the *network structure* is concerned, a delabeled graph model is equivalent to its labeled source, while both are different from the corresponding unlabeled model as Fig. 2 illustrates.

The key points of this illustrative section motivating what follows are as follows:

- (i) exchangeability is indistinguishability of random variables;
- (ii) therefore, models of real networks with distinguishable nodes can be either nonexchangeable or unlabeled;
- (iii) the correct null models of the structure of real networks with distinguishable nodes must be models of unlabeled graphs;
- (iv) the statistical properties of unlabeled graph models can be very different from their labeled and delabeled counterparts;
- (v) compared to their labeled and delabeled counterparts, unlabeled graph models have been studied much more poorly because they are much more difficult to deal with, see Sec. V.

III. ERDŐS-RÉNYI GRAPHS AND CONFIGURATION MODEL

A. Erdős-Rényi graphs

As discussed in the previous section, the microcanonical labeled and unlabeled Erdős-Rényi (ER) graph models $\mathcal{G}_{n,m}$ and $\mathcal{U}_{n,m}$ are *defined* by the entropy-maximizing uniform probability distributions $P(G) = 1/|\mathcal{G}_{n,m}|$ and $P(U) = 1/|\mathcal{U}_{n,m}|$ over all labeled graphs $G \in \mathcal{G}_{n,m}$ and, respectively, unlabeled graphs $U \in \mathcal{U}_{n,m}$ with n nodes and m edges. While the number of labeled graphs with n nodes and m edges is exactly $|\mathcal{G}_{n,m}| = \binom{N}{m}$ where $N = \binom{n}{2}$, the number of unlabeled graphs with n nodes and m edges $|\mathcal{U}_{n,m}|$ is known only asymptotically for large graphs [41].

The conjugated canonical versions of $\mathcal{G}_{n,m}$ and $\mathcal{U}_{n,m}$ are $\mathcal{G}_{n,p}$ and $\mathcal{U}_{n,p}$. These are the maximum-entropy labeled and unlabeled graphs of size n in which the number of links is not fixed exactly to m ; instead, the average number of links \bar{m} is fixed to pN , or equivalently, the average graph density $\bar{d} = \bar{m}/N$ is fixed to p . While $\mathcal{G}_{n,p}$ is as well studied as $\mathcal{G}_{n,m}$, the unlabeled canonical ER graphs $\mathcal{U}_{n,p}$ has never been considered before, so we *define* them next, after recalling the basic entropic facts about $\mathcal{G}_{n,p}$.

1. Unlabeled canonical ER graphs $\mathcal{U}_{n,p}$

As is well known, a $\mathcal{G}_{n,p}$ graph G can be generated by linking all pairs of labeled nodes independently with probability p . The resulting probability to generate graph G in the model is

$$P(G) = p^{m(G)}(1-p)^{N-m(G)}, \quad (1)$$

where $m(G)$ is the number of edges in G . This probability distribution is the canonical entropy-maximizing Gibbs (a.k.a. exponential family) distribution, since it can be rewritten in the Gibbs form [1]

$$P(G) = \frac{\exp[\beta m(G)]}{Z}, \quad (2)$$

where the partition function involves the summation over all labeled graphs \mathcal{G}_n of size n ,

$$Z = \sum_{G \in \mathcal{G}_n} \exp[\beta m(G)], \quad (3)$$

which can be shown simplifies to

$$Z = (e^\beta + 1)^N. \quad (4)$$

The inverse temperature parameter $\beta \in \mathbb{R}$ is related to $p \in [0, 1]$ via

$$p = \frac{1}{e^{-\beta} + 1}, \quad (5)$$

which is the solution of the standard free-energy equation

$$\frac{\partial \log Z}{\partial \beta} = \sum_{G \in \mathcal{G}_n} m(G)P(G) = \bar{m} = pN. \quad (6)$$

The unlabeled ER graphs U in $\mathcal{U}_{n,p}$ are thus *defined* by the entropy-maximizing probability distribution of the same

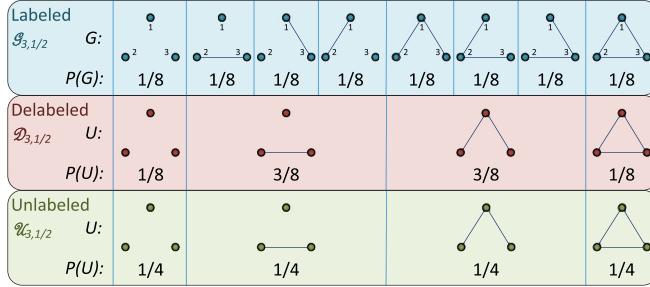


FIG. 4. The labeled, delabeled, and unlabeled canonical Erdős-Rényi random graphs of size $n = 3$ and average density $p = 1/2$. The probabilities of all graphs in the models are shown at the bottom of each row. The entropies of the three models are $S_L = 3$, $S_U = 2$, and $S_D \approx 1.81$ bits, reflecting the general inequality $S_D \leq S_U \leq S_L$.

Gibbs form,

$$P(U) = \frac{\exp[\beta m(U)]}{Z}, \quad (7)$$

except that the graphs are unlabeled, so the partition function involves the summation not over all the n -sized *labeled* graphs \mathcal{G}_n but over the much smaller but also much more intractable space \mathcal{U}_n of all the *unlabeled* graphs of size n ,

$$Z = \sum_{U \in \mathcal{U}_n} \exp[\beta m(U)]. \quad (8)$$

Unfortunately, this sum does not in general simplify to anything as nice looking as (4). As a consequence, there is no nice-looking analogy of (1) for $P(U)$, which, among many other things, implies that the $\mathcal{U}_{n,p}$ graphs *cannot* be generated by placing edges independently with probability p among N unlabeled node pairs. By doing so, you generate an unlabeled graph not from the unlabeled model $\mathcal{U}_{n,p}$ but from the delabeled one $\mathcal{D}_{n,p}$. The free-energy equation (6) linking β to p holds,

$$\frac{\partial \log Z}{\partial \beta} = \sum_{U \in \mathcal{U}_n} m(U)P(U) = \bar{m} = pN, \quad (9)$$

but does not lead to anything as simple as (5). Yet we can show that the solution of (9) exists and is unique for any n, p .

Using graph complementarity arguments, we can also show that the solution of (9) with $p = 1/2$ yields $\beta = 0$ (infinite temperature) for any n , resulting in the uniform distribution over all the unlabeled graphs \mathcal{U}_n of size n as Fig. 4 illustrates for the simplest nontrivial case $n = 3$. In $\mathcal{U}_{3,p}$, the direct evaluation of (8) yields the partition function

$$Z = \frac{e^{4\beta} - 1}{e^{\beta} - 1}, \quad (10)$$

so the probability of the four unlabeled graphs of size 3 with $m = 0, 1, 2, 3$ edges is $P(m) = e^{m\beta}/Z$, while (9) leads to

$$P = \frac{1}{3} \left[\frac{1}{e^{-\beta} + 1} + \tanh \beta + 1 \right]. \quad (11)$$

Figure 4 also illustrates the key statistical differences between the labeled and unlabeled ER graphs. For instance, the probability to generate a graph with $m = 0, 1, 2, 3$ edges in the

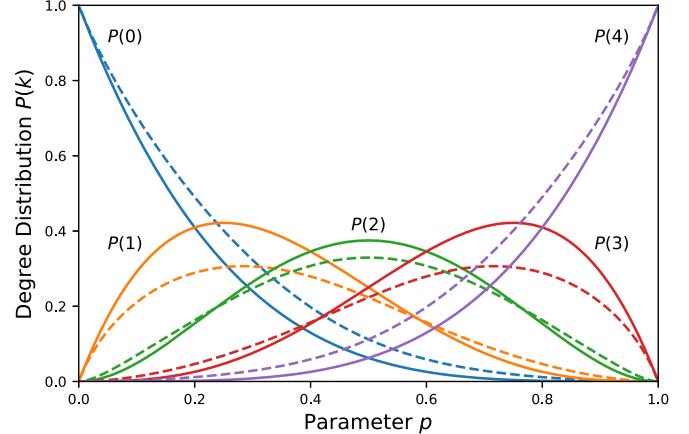


FIG. 5. The degree distributions in the canonical labeled $\mathcal{G}_{n,p}$ and unlabeled $\mathcal{U}_{n,p}$ ER graphs of size $n = 5$. The solid and dashed curves show the exact solutions for the probability $P(k)$ that a random node in a random labeled $\mathcal{G}_{5,p}$ graph and, respectively, unlabeled $\mathcal{U}_{5,p}$ graph has degree k for all values of degree $k = 0, 1, 2, 3, 4$ and density $p \in [0, 1]$.

labeled and delabeled cases with $\beta = 0$ is given by the binomial distribution $1/8, 3/8, 3/8, 1/8$, while in the unlabeled case this probability is uniform, $1/4$ for any m . The statistical similarities between the delabeled and labeled graphs, both different from the unlabeled ones, are similar to those in the microcanonical case in Fig. 2.

2. Degree distribution

Another key difference between (de)labeled and unlabeled ER graphs is the degree distribution. As can be deduced from Fig. 4, the degree distribution in the unlabeled graphs is uniform, $P(k) = 1/3$ for degrees $k = 0, 1, 2$, versus the corresponding binomial distribution $1/4, 1/2, 1/4$ in the labeled graphs. Figure 5 shows the exact degree distributions in the canonical labeled and unlabeled ER graphs $\mathcal{G}_{n,p}$ and $\mathcal{U}_{n,p}$ of size $n = 5$. We see that they are different for any values of $p \neq \{0, 1\}$. The unlabeled graphs always have more nodes of degree 0, for instance.

The degree distribution in $\mathcal{U}_{n,p}$ is unknown, and we leave it as an open problem to compare it against the degree distribution in $\mathcal{U}_{n,m}$, which, as was shown in Ref. [42], is very different from the Poisson one in sparse $\mathcal{G}_{n,m}$. In particular, if $m = \bar{k}n/2$ with constant \bar{k} , then the $\mathcal{U}_{n,m}$ graph consists of a connected component of size

$$\ell \approx \frac{2m}{\log m}, \quad (12)$$

the average degree and degree distribution in which are

$$\bar{k}_{>0} = \frac{2m}{\ell} \approx \log m \approx \log \ell \approx \log n, \quad (13)$$

$$P_\ell(k) \approx \frac{\log^k \ell}{\ell k!}, \quad k > 0. \quad (14)$$

However, most nodes are not in this component and have degree 0; their number is $n - \ell \approx n$. The graph is thus dominated by isolated nodes. However, if they are ignored, then it

has a Poisson-like degree distribution $P_\ell(k)$ with a logarithmically diverging average degree.

3. Entropy

Notwithstanding these drastic structural differences reflected in the degree distribution, the leading terms of entropy of $\mathcal{G}_{n,m}$ and $\mathcal{U}_{n,m}$ are surprisingly the same. As can be deduced from Ref. [41], the entropy of $\mathcal{U}_{n,m}$ is

$$S_U = \frac{\bar{k}}{2}n \log n - \bar{k}n \log \log n + \frac{\bar{k}}{2}(\log \bar{k} - 1)n + o(n) \quad (15)$$

for $\bar{k} \ll \log n$. While the leading term is the same, the subleading terms are different than in the labeled $\mathcal{G}_{n,m}$ whose entropy for $\bar{k} \ll n$ is [5]

$$S_L = \frac{\bar{k}}{2}n \log n - \frac{\bar{k}}{2}(\log \bar{k} - 1)n + o(\bar{k}n), \quad (16)$$

so $S_U < S_L$ for sufficiently large n . Due to the connectivity phase transition at $\bar{k} \sim \log n$, the $\mathcal{U}_{n,m}$ graphs with $\bar{k} \gg \log n$ do not have any degree-0 nodes and consist of a single connected component, which is asymmetric [42]. Since the graph is asymmetric, the labeled and unlabeled entropies are related by $S_U = S_L - \log n!$ in this denser case, but since $\bar{k} \gg \log n$, this difference is negligible.

Nothing is known about the entropy of the canonical $\mathcal{U}_{n,p}$ model in any regime, including whether there is any ensemble equivalence between $\mathcal{U}_{n,p}$ and $\mathcal{U}_{n,m}$ akin to the one established for the labeled $\mathcal{G}_{n,p}$ and $\mathcal{G}_{n,m}$ [14,15]. We leave these as open problems as well.

B. Configuration model

While the entropy of the labeled microcanonical configuration model (CM) is a well-explored subject [2–5,43–45], very little is known about its unlabeled version. The existing results [46,47] tell only whether the CM graphs are asymmetric or symmetric, i.e., whether they have *any* nontrivial automorphisms, not *how many* automorphisms they have, as is needed for entropy calculations.

As far as sparse power-law degree sequences with exponent γ are concerned, the latest results in Ref. [47] show that if $\gamma > 3$, then the CM graphs are symmetric. However, it remains unknown what happens for $\gamma \leq 3$. Since the key ingredients that break asymmetry (“most graphs are asymmetric” [48]) are hubs in tandem with low-degree nodes (star graphs are “very symmetric”), the proofs of graph asymmetry (leading to $S_U = S_L - \log n!$) involve strict bounds on the maximum degree and the numbers of nodes of degree 1 and 2 [46,47], which are violated in sparse scale-free degree sequences with $\gamma \leq 3$. However, it is still not excluded that such graphs are asymmetric or, much more likely, that the number of their automorphisms is $\ll n!$. If so, then their unlabeled entropy would be $S_U \approx (\bar{k}/2 - 1)n \log n$ since $S_L \approx (\bar{k}/2)n \log n$ [4]. (Dis)proving this is yet another open problem.

Similarly to the unlabeled canonical ER $\mathcal{U}_{n,p}$, the unlabeled canonical CM has not been mentioned in the past, so we *define* it here. We first recall that the labeled canonical CM, or the soft configuration model (SCM) [1,2,7], is defined by a sequence of expected degrees $\{\kappa_i\}$ of nodes $i \in [n]$. The Gibbs

probability distribution of random labeled graphs in the model is

$$P(G) = \frac{\exp \left[\sum_{i=1}^n \beta_i d_i(G) \right]}{Z}, \quad (17)$$

where $d_i(G)$ is the degree of node i in graph $G \in \mathcal{G}_n$, and the parameters $\{\beta_i\}$ are found as the solution of the system of n free-energy equations,

$$\frac{\partial \log Z}{\partial \beta_i} = \sum_{j=1}^n p_{ij} = \kappa_i, \quad (18)$$

where

$$p_{ij} = \frac{1}{e^{-\beta_i - \beta_j} + 1} \quad (19)$$

are the probabilities of edges between nodes i and j . The model is *not* exchangeable, unless all κ_i s are the same.

This vanilla SCM definition is clearly not directly applicable to unlabeled graphs since it explicitly refers to node labels i via κ_i . However, the following alternative SCM definition based on the empirical degree distribution in graph $G \in \mathcal{G}_n$ avoids this problem:

$$P(G) = \frac{\exp \left[\sum_{k=0}^{n-1} \alpha_k n_k(G) \right]}{Z}. \quad (20)$$

Here $n_k(G)$ is the number of nodes of degree k in G . This version of the labeled SCM is exchangeable, and its *definition* is directly applicable to unlabeled graphs $U \in \mathcal{U}_n$:

$$P(U) = \frac{\exp \left[\sum_{k=0}^{n-1} \alpha_k n_k(U) \right]}{Z}. \quad (21)$$

As in the ER case, the main difference between the labeled and unlabeled SCMs defined in Eqs. (20) and (21) is that the partition function Z involves the summation over all labeled versus unlabeled graphs of size n . Both models are defined by desired expected numbers $\{\nu_k\}$ of nodes of degree k , and the parameters $\{\alpha_k\}$ are the solutions of the standard free-energy equations,

$$\frac{\partial \log Z}{\partial \alpha_k} = \nu_k. \quad (22)$$

IV. RANDOM GEOMETRIC GRAPHS

Calculating the entropy of random geometric graphs (RGGs) is a cornerstone problem in estimating entropy of a large class of labeled network models with hidden variables [49–52], where the connection probability between nodes i and j is $p_{ij} = p(x_i, x_j)$, where $p(x, y) \in [0, 1]$ is a function of i, j ’s random coordinates x_i, x_j in some space. These models include not only all spatial networks and latent space models but also the soft versions of the configuration model, preferential attachment, stochastic block model, as well as graphons in dense graphs. In dense graphs, the number of edges is $m \sim n^2$, so that their entropy, known as graphon entropy [53], always dominates the coordinate entropy coming from random x_i ’s. However, in sparse graphs, the numbers of edges and node coordinates are of the same order $\sim n$, so their entropies may be comparable. In RGGs, the graph entropy is exactly the coordinate entropy, since given

the coordinates, edges exist deterministically. Therefore, the estimation of the RGG entropy is of utmost importance for disentangling the edge and coordinate entropies.

Here we focus on the sparse one-dimensional RGGs *defined* by

(1) sprinkling n points uniformly at random over the interval $[0, n]$, and then

(2) linking all pairs of points at distance $< r$ on $[0, n]$, where $r > 0$ is a constant,

so the expected average degree converges to $\bar{k} = 2r$. Step (1) implements the binomial point process of rate 1 on $[0, n]$, while step (2) says that $p(x, y) = \mathbb{1}\{|x - y| < r\}$, where $\mathbb{1}$ is the indicator function.

Observe that as defined above, the graphs are actually unlabeled because we did not label the sprinkled points. This is consistent with the general definition of a point process in probability as a random point measure [54], which does not involve any labeling. Denote the entropy of the resulting unlabeled RGGs by S_U .

We can also modify step (1) in the definition of unlabeled RGGs above to the following:

(1') sample the coordinates x_i of nodes $i \in [n]$ from the uniform distribution on $[0, n]$ i.i.d.'ly.

The points are now labeled by integers $i \in [n]$, so the resulting graph is labeled as well, but its labels are completely random. That is, it is easy to see that generating the labeled graph using steps (1', 2) is equivalent to generating the unlabeled graph using steps (1,2) first and then labeling it by one out of the $n!$ possible permutations of labels $[n]$ selected uniformly at random. Denote the entropy of the resulting labeled RGGs by S_L .

Going back from the labeled to unlabeled graphs is achieved by generating a labeled graph and then delabeling it. This means that in contrast with ER, the unlabeled and delabeled models of RGGs are actually identical. This is not a surprise but a reflection of the general situation: If a network model is a maximum-entropy null model, as is the case with ER, then its unlabeled and delabeled versions are usually different. However, if a network model is defined by a graph-generation process in which labels do not matter, as is the case with RGGs, then the unlabeled and delabeled versions are identical.

Unfortunately, the entropy of neither unlabeled RGGs S_U nor labeled ones S_L is amenable to any brute-force calculations due to the intractable dependencies among edges, so we need to devise some tricks, which are described in the following sections.

A. An upper bound for unlabeled entropy

First, we upper bound the unlabeled entropy S_U by the entropy of the uniform distribution over all unlabeled graphs that can be realized as one-dimensional geometric graphs, i.e., unit interval graphs. This entropy is $\log \mathcal{N}_U$, where \mathcal{N}_U is the number of such graphs, which is

$$\mathcal{N}_U = \frac{4^n}{c\sqrt{\pi n^3}}, \quad (23)$$

where c is approximately 5.01 [55]. As a side note, the number of orderly labeled geometric graphs with $x_1 < x_2 <$

$\dots < x_n$, whose entropy S_O is squeezed between S_U and S_L , $S_U \leq S_O \leq S_L$, is the Catalan number (Exercise 6.19 in Ref. [56]),

$$\mathcal{N}_O = \frac{\binom{2n}{n}}{n+1} \approx \frac{4^n}{\sqrt{\pi n^3}}. \quad (24)$$

Therefore, our first result is that the entropy of unlabeled RGGs is

$$S_U \leq \log \mathcal{N}_U \approx n \log 4. \quad (25)$$

The application of the same technology to S_L would tell us that $S_L \lesssim n \log n$, since $\log \mathcal{N}_L \sim n \log n$ [55], but it would not lead to any lower bound for S_L , so it could still be that $S_L \sim n \sim S_U$. We derive much tighter upper and lower bounds for S_L using a different route.

B. An upper bound for labeled entropy

We first recall a very simple and general relation between the labeled and delabeled entropies [39,40]. We call the latter the unlabeled entropy below, since unlabeled RGGs are identical to delabeled RGGs.

Consider any model of labeled graphs $G \in \mathcal{G}_n$ of size n with distribution $P(G)$ whose entropy is

$$S_L = - \sum_{G \in \mathcal{G}_n} P(G) \log P(G). \quad (26)$$

Let U_G be the unlabeled version of G , and let \mathcal{G}_U be the isomorphism class corresponding to an unlabeled graph $U \in \mathcal{U}_n$:

$$\mathcal{G}_U = \{G \in \mathcal{G}_n : U_G = U\}. \quad (27)$$

Denote its size by $\mathcal{N}_U = |\mathcal{G}_U|$ and observe that

$$\mathcal{N}_U = \frac{n!}{|\text{Aut}(U)|}, \quad (28)$$

where $\text{Aut}(U)$ is the group of automorphisms of any labeled version of U . Let $P(U)$ be the delabeled probability distribution induced by $P(G)$,

$$P(U) = \sum_{G \in \mathcal{G}_U} P(G). \quad (29)$$

Observe that since any labeled graph G has only one unlabeled graph U_G corresponding to it, we have that

$$P(G) = \sum_{U \in \mathcal{U}_n} P(G|U)P(U) = P(G|U_G)P(U_G). \quad (30)$$

From here it follows that the conditional distribution of graph $G \in \mathcal{G}_U$ given that its unlabeled graph is U_G is

$$P(G|U_G) = \frac{P(G)}{P(U_G)}. \quad (31)$$

If the model is exchangeable, as is the case with labeled RGGs, then $P(G|U_G)$ is uniform,

$$P(G|U_G) = \frac{1}{\mathcal{N}_{U_G}}, \quad (32)$$

hence

$$P(G) = \frac{P(U_G)}{\mathcal{N}_{U_G}}. \quad (33)$$

Substituting this into S_L yields

$$\begin{aligned} S_L &= - \sum_{G \in \mathcal{G}_n} P(G) \log \left[\frac{P(U_G)}{\mathcal{N}_{U_G}} \right] \\ &= \log n! - \sum_{U \in \mathcal{U}_n} \sum_{G \in \mathcal{G}_U} P(G) [\log P(U) + \log |\text{Aut}(U)|] \\ &= \log n! - \sum_{U \in \mathcal{U}_n} P(U) \log P(U) \\ &\quad - \sum_{U \in \mathcal{U}_n} P(U) \log |\text{Aut}(U)|. \end{aligned} \quad (34)$$

That is,

$$S_L = S_U + \log n! - \mathcal{A}, \quad (35)$$

where

$$S_U = - \sum_{U \in \mathcal{U}_n} P(U) \log P(U) \quad (36)$$

is the unlabeled entropy and

$$\mathcal{A} = \langle \log |\text{Aut}(U)| \rangle = \sum_{U \in \mathcal{U}_n} P(U) \log |\text{Aut}(U)| \quad (37)$$

is the expected logsize of the automorphism group.

Equation (35) provides the following useful upper and lower bounds for the labeled entropy:

$$S_L \leq S_U + \log n!, \quad (38)$$

$$S_L \geq \log n! - \mathcal{A}. \quad (39)$$

Since $S_U \lesssim n$ in our RGGs, we immediately arrive at the upper bound for their labeled entropy using (38):

$$S_L \leq n \log n. \quad (40)$$

An upper bound on \mathcal{A} would yield a lower bound on S_L using (39), which we deal with next.

C. A lower bound for labeled entropy

We first assume that n is sufficiently large, so we can approximate the binomial point process of rate 1 on $[0, n]$ with the Poisson one, where the distances d between consecutive points are independent exponentially distributed random variables with PDF $P(d) = e^{-d}$ [57]. We then recall that the percolation threshold in one-dimensional RGGs is infinite, simply because $d > r$ with probability $p = e^{-r}$. It follows that the sizes s_c of connected components $c \in [C]$ are independent geometrically distributed random variables with PDF $P(s) = p(1 - p)^{s-1}$, while the number of components C is approximately binomial, $P(C) = \binom{n-1}{C-1} p^{C-1} (1 - p)^{n-C}$, $C \in [n]$.

The key observation then is that for a label permutation to be an automorphism, it must permute nodes within a component, permute the components, or both. A trivial upper bound

$$A = C! \prod_{c=1}^C s_c! \quad (41)$$

for the number of automorphisms is when all components c are maximally symmetric, i.e., when they all are complete graphs of size s_c . It follows that

$$\log A \approx C \log C - C + \sum_{c=1}^C \log s_c!, \quad (42)$$

and since $\langle C \rangle \approx pn$ and $\langle s_c \rangle = 1/p = e^r$, we see that the leading term in $\langle \log A \rangle$ is $\langle C \log C \rangle$, which one can check is $\approx pn \log n$. We thus have that

$$\mathcal{A} \leq pn \log n. \quad (43)$$

Substituting this into (39), and combining with the upper bound obtained earlier, we finally get

$$(1 - e^{-k/2})n \log n \leq S_L \leq n \log n. \quad (44)$$

We note that the larger the average degree, the tighter these bounds, although they are asymptotic, holding for $\langle C \rangle \gg 1$, meaning $n \gg e^{k/2}$.

Comparing (44) with (25), we conclude that $S_U \ll S_L$. That is, the entropy of random labeling $\log n!$ in (38) and (39) dominates the network-structural entropy (25).

V. CONCLUSIONS

Null models of networks are typically used to assess the statistical significance of network features in a given real-world network and to investigate whether such features are relevant for a particular set of network functions [58]. Colloquially, null models are “maximally random” models of networks constrained to have a set of particular network properties. Real networks are compared against their null models to detect statistically significant deviations in the values of network properties that are not constrained in the null model. The “maximum randomness” of null models is formally achieved by maximizing the model entropy under the imposed constraints. But should these models be models of labeled or unlabeled networks? This elephant-in-the-room question has been magically ignored in the past research.

Here we began with a crucial observation that if nodes in a real-world network are labeled and *distinguishable*, then correct null models for such a network must be either *nonexchangeable* or *unlabeled*, simply because exchangeability is statistical indistinguishability. Exchangeable models represent a world which is statistically very different from any data source generating distinguishable labels. And if we are interested only in the network structure, then the correct null models of the structure of real networks with distinguishable nodes must be models of unlabeled graphs.

The logical next two questions are then as follows:

(1) Does it really matter what models to consider, labeled or unlabeled, as they may be equivalent upon some simple transformation?

(2) If it does matter, and unlabeled models should have been studied all these years as much as—if not more than—labeled ones, then why has this not been the case?

To address the first question, we have shown here that labeled and unlabeled models of sparse networks can be very different and nonequivalent. The emphasis on *sparse* networks

is important here. Almost all *dense* networks are asymmetric [48]. Therefore, there exists a simple relation between their labeled and unlabeled models: Almost every unlabeled network corresponds to exactly $n!$ labeled ones with all possible permutations of labels. It follows then that as far as entropy is concerned, for instance, the general equation (35) relating the labeled and unlabeled network entropies S_L and S_U becomes trivial: $S_L \approx S_U + \log n!$. But since the unlabeled entropy of dense networks is $S_U \sim n^2$ [13,53], this difference between S_L and S_U is negligible, $S_L \sim S_U \sim n^2$. In other words, in dense networks, there is room for huge diversity of network structure. Therefore, randomness associated with network structure, represented by the unlabeled entropy $S_U \sim n^2$, dominates the labeling entropy $S_L = S_L - S_U \approx n \log n$.

We have shown that in sparse networks, the situation is very different. The sources of network-structural and labeling randomnesses are comparable in their power, so the network-structural entropy S_U may or may not be the leading factor, depending on the model. We have considered three examples demonstrating all of the three possibilities:

(i) in Erdős-Rényi graphs, the network-structural entropy dominates the labeling entropy:

$$S_U \gg S_L;$$

(ii) in the scale-free configuration model, the two entropies are conjecturally comparable:

$$S_U \sim S_L;$$

(iii) in random geometric graphs, the labeling entropy dominates the network-structural entropy:

$$S_U \ll S_L.$$

Even though the network-structural entropy wins in sparse Erdős-Rényi graphs, this example is still a source of concerns for at least two reasons. First, by definition, Erdős-Rényi graphs are maximally random graphs with a given average degree, so this is where we could expect to see the strongest domination of S_U over S_L compared to other sparse network models. However, this domination is not really strong. In fact, it is marginal: As can be seen from (15) and (16), $S_U/S_L \approx (1/2) \log n / \log \log n$.

Worse, the entropic equivalence is definitely a necessary but not sufficient condition for model equivalence. The Erdős-Rényi example shows that even though the labeled and unlabeled models are entropically equivalent ($S_L \approx S_U$), their very basic structural property—the degree distribution—is very different between the two models.

Compared to Erdős-Rényi graphs, the random geometric graph example is much more disconcerting, as it clearly demonstrates that even the entropic equivalence between labeled and unlabeled models can be broken in sparse networks and that the entropy of meaningless labeling noise S_L may easily be the leading factor, overpowering the network-structural signal S_U . This result is disconcerting because it implies that entropy maximization in sparse networks may easily be the maximization of the entropy of meaningless labeling that you do not care about versus the intended maximization of the entropy of the network structure. Such caveats can easily lead to profound aberrations and statistical errors in conclusions made about the structure of sparse real-world networks based on their maximum-entropy labeled models.

It is important to note that in some situations, working with labeled network models is justified. As mentioned in Sec. II,

models of real-world networks with indistinguishable nodes must be labeled. Other examples include situations where labeling is statistically meaningful. One such example, briefly mentioned in Sec. IV, is the random geometric graph labeled in the order of increasing coordinates. Thus labeled, random geometric graphs can be formulated as a growing network model [59]. This example generalizes to any growing network model, such as preferential attachment whose symmetry properties were considered in Ref. [60], where the preferred labeling is by nodes' birth times. Yet another example of a different sort is the stochastic block model where nodes can be labeled by communities they belong to [21,22].

Unfortunately, even in cases where there exists a (unique) preferred labeling scheme, the models that have been actually studied at depth in the past are the models of networks labeled by $[n]$ arbitrarily. Even more unfortunate is that this practice is nearly never spelt out explicitly. As a rule of thumb, if unsure what networks, labeled or unlabeled, a particular result in network science or graph theory is about, then assume it is about arbitrarily labeled networks.

Perhaps the *main implication* of the results in this paper is that this practice of silence about the differences between labeled and unlabeled network models should be abandoned, at least as far as applications to sparse networks are concerned. It is really imperative to understand in what situations the replacement of the correct representation of the network structure, an unlabeled network, by its simpler labeled surrogate, is statistically justified. As a bare minimum, it should always be made clear what networks we are dealing with, labeled or unlabeled, and why. As we have seen here, the answer to this *why* question may be very difficult. It may not even exist, suggesting a dire need for a thorough reexamination of the foundational results in the science of sparse networks. To motivate this reexamination, note that its very basic starting point, the degree distribution in unlabeled Erdős-Rényi graphs, is not what you would have expected.

Given that the very core of network science is all about the *structure* (and function) of complex networks [61], one may ask why all of the best-studied null models of the structure of networks are labeled, not unlabeled, as they should have been. Why did unlabeled graphs not attract the deserved attention even in textbooks on network science and graph theory? We speculate that the main reason is that it is actually quite difficult not only to think about unlabeled networks but also to deal with them in practice.

First, it is quite a challenge to store an unlabeled graph on a computer. The standard practice to represent an unlabeled graph in textbooks, web sites, or computer programs like *Mathematica* is to picture it. Yet storing graphs in pictures does not get you far computationally, so *Mathematica* and similar programs rely on *canonical labeling* [62] of unlabeled graphs, an important area of research in graph theory and computer science looking for computationally efficient ways to assign a unique labeling to an unlabeled graph.

This “mundane” graph representation issue is a nuance, compared to the problem of graph generation. It is quite a challenge to generate even the simplest unlabeled graphs, i.e., unlabeled Erdős-Rényi graphs. Some of the best results on how to generate microcanonical unlabeled Erdős-Rényi graphs are available in Ref. [63]. Since the canonical

unlabeled Erdős-Rényi graphs have not been considered or even mentioned before, there are no results whatsoever on how to generate them. Nor are we aware of any results on how to generate graphs in the unlabeled configuration model, either microcanonical or canonical, the latter also defined for the first time in this paper only.

Yet it is not the case that there are no strongly positive results on unlabeled graphs at all. The optimistic results in Ref. [42] say that in microcanonical unlabeled Erdős-Rényi graphs, the values of a huge class of network properties can be linked to their values in the corresponding labeled graphs. Roughly, the unlabeled values are the labeled values in graphs of a different effective size. Unfortunately, these results apply mostly to denser graphs with the average degree $\bar{k} \gg \log n$.

Another source of optimism is that sufficient statistics in exchangeable models are unlabeled properties. By *unlabeled properties*, we mean those network properties whose values are identical across all labeled graphs in any isomorphism class. Therefore, there exists a unique unlabeled model corresponding to each exchangeable model, and vice versa, having identical sufficient statistics, differing only in their sample spaces, i.e., unlabeled vs. labeled graphs. As discussed in Sec. II, if all of the network properties that a researcher has data for and wants to model in a null model are unlabeled, then a natural setting for the sample space would be that of unlabeled graphs. Regardless of the choice of an unlabeled

versus exchangeably labeled model in such cases, the discussed connection between these two types of models may help to translate ideas, methods, and techniques between the two settings.

To end on a truly positive note, the last remark is that unlabeled network models completely avoid the exchangeability conundrum in sparse networks. On the one hand, if node labels are meaningless integers that “do not matter,” then the network model must be exchangeable since the probability of a network in the model cannot depend on how the network is labeled. On the other hand, the thermodynamic limit of any sparse exchangeable network model necessarily consists of empty networks due to the Aldous-Hoover theorem [64,65]. Exchangeability makes no sense in the realm of unlabeled graphs, so the paradox dissolves there. The Aldous-Hoover theorem simply implies that the limits of sparse unlabeled networks—whatever they are—cannot be exchangeably labeled by integers.

ACKNOWLEDGMENTS

We thank O. Kallenberg, Y. Peres, M. Rácz, Z. Burda, P. Krapivsky, N. Wormald, T. Łuczak, F. Radicchi, and G. Bianconi for useful discussions and suggestions. This work was supported by the NSF Grant No. IIS-1741355.

- [1] J. Park and M. E. J. Newman, Statistical mechanics of networks, *Phys. Rev. E* **70**, 066117 (2004).
- [2] G. Bianconi, The entropy of randomized network ensembles, *Europhys. Lett.* **81**, 28005 (2008).
- [3] G. Bianconi, A. C. C. Coolen, and C. Perez Vicente, Entropies of complex networks with hierarchically constrained topologies, *Phys. Rev. E* **78**, 016114 (2008).
- [4] G. Bianconi, Entropy of network ensembles, *Phys. Rev. E* **79**, 036114 (2009).
- [5] K. Anand and G. Bianconi, Entropy measures for networks: Toward an information theory of complex topologies, *Phys. Rev. E* **80**, 045102 (2009).
- [6] K. Anand, G. Bianconi, and S. Severini, Shannon and von Neumann entropy of random networks with heterogeneous expected degree, *Phys. Rev. E* **83**, 036109 (2011).
- [7] D. Garlaschelli and M. Loffredo, Maximum likelihood: Extracting unbiased information from complex networks, *Phys. Rev. E* **78**, 015101 (2008).
- [8] D. Garlaschelli and M. Loffredo, Generalized Bose-Fermi Statistics and Structural Correlations in Weighted Networks, *Phys. Rev. Lett.* **102**, 038701 (2009).
- [9] T. Squartini and D. Garlaschelli, Analytical maximum-likelihood method to detect patterns in real networks, *New J. Phys.* **13**, 083001 (2011).
- [10] R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli, Enhanced reconstruction of weighted networks from strengths and degrees, *New J. Phys.* **16**, 043022 (2014).
- [11] T. Squartini, R. Mastrandrea, and D. Garlaschelli, Unbiased sampling of network ensembles, *New J. Phys.* **17**, 023052 (2015).
- [12] K. Zuev, O. Eisenberg, and D. Krioukov, Exponential random simplicial complexes, *J. Phys. A: Math. Theor.* **48**, 465002 (2015).
- [13] P. van der Hoorn, G. Lippner, and D. Krioukov, Sparse maximum-entropy random graphs with a given power-law degree distribution, *J. Stat. Phys.* **173**, 806 (2018).
- [14] K. Anand and G. Bianconi, Gibbs entropy of network ensembles by cavity methods, *Phys. Rev. E* **82**, 011116 (2010).
- [15] T. Squartini, J. de Mol, F. den Hollander, and D. Garlaschelli, Breaking of Ensemble Equivalence in Networks, *Phys. Rev. Lett.* **115**, 268701 (2015).
- [16] D. Garlaschelli, F. den Hollander, and A. Roccaverde, Ensemble nonequivalence in random graphs with modular structure, *J. Phys. A: Math. Theor.* **50**, 015001 (2017).
- [17] C. E. Shannon, A mathematical theory of communication, *Bell. Syst. Tech. J.* **27**, 379 (1948).
- [18] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, Hoboken, NJ, 2005).
- [19] G. Bianconi, P. Pin, and M. Marsili, Assessing the relevance of node features for network structure, *Proc. Natl. Acad. Sci. USA* **106**, 11433 (2009).
- [20] K. Zhao, A. Halu, S. Severini, and G. Bianconi, Entropy rate of nonequilibrium growing networks, *Phys. Rev. E* **84**, 066113 (2011).
- [21] T. P. Peixoto, Entropy of stochastic blockmodel ensembles, *Phys. Rev. E* **85**, 056122 (2012).
- [22] T. P. Peixoto, Parsimonious Module Inference in Large Networks, *Phys. Rev. Lett.* **110**, 148701 (2013).
- [23] K. Anand, D. Krioukov, and G. Bianconi, Entropy distribution and condensation in random networks with a given degree distribution, *Phys. Rev. E* **89**, 062807 (2014).

[24] M. Z. Rácz and S. Bubeck, Basic models and questions in statistical network analysis, *Stat Surv* **11**, 1 (2017).

[25] F. Radicchi and C. Castellano, Uncertainty Reduction for Stochastic Processes on Complex Networks, *Phys. Rev. Lett.* **120**, 198301 (2018).

[26] F. Radicchi, D. Krioukov, H. Hartle, and G. Bianconi, Classical information theory of networks, *J. Phys. Complex.* **1**, 025001 (2020).

[27] G. Bianconi, Statistical physics of exchangeable sparse simple networks, multiplex networks, and simplicial complexes, *Phys. Rev. E* **105**, 034310 (2022).

[28] G. Bianconi, Grand canonical ensembles of sparse networks and Bayesian inference, *Entropy* **24**, 633 (2022).

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

[30] A. Coutrot, E. Manley, S. Goodroe, C. Gahnstrom, G. Filomena, D. Yesiltepe, R. C. Dalton, J. M. Wiener, C. Hölscher, M. Hornberger, and H. J. Spiers, Entropy of city street networks linked to future spatial navigation ability, *Nature (London)* **604**, 104 (2022).

[31] J. P. Coon, C. P. Dettmann, and O. Georgiou, Entropy of spatial network ensembles, *Phys. Rev. E* **97**, 042319 (2018).

[32] M.-A. Badiu and J. P. Coon, in *Proceedings of the IEEE International Symposium on Information Theory* (IEEE, Vail, CO, 2018), pp. 2137–2141.

[33] S. Bubeck, J. Ding, R. Eldan, and M. Z. Rácz, Testing for high-dimensional geometry in random graphs, *Random Struct. Alg.* **49**, 503 (2016).

[34] S. Liu and M. Z. Rácz, Phase transition in noisy high-dimensional random geometric graphs (2021), [arXiv:2103.15249](https://arxiv.org/abs/2103.15249).

[35] F. Harary and E. M. Palmer, *Graphical Enumeration* (Academic Press, New York, 1973).

[36] L. Papadopoulos, M. A. Porter, K. E. Daniels, and D. S. Bassett, Network analysis of particles and grains, *J. Complex Netw.* **6**, 485 (2018).

[37] C. Kittel and H. Kroemer, *Thermal Physics*, 2nd ed. (W. H. Freeman & Company, New York, 1980).

[38] I. Ford, *Statistical Physics: An Entropic Approach* (John Wiley & Sons, Ltd, Chichester, UK, 2013).

[39] Y. Choi and W. Szpankowski, Compression of graphical structures: Fundamental limits, algorithms, and experiments, *IEEE Trans Inf Theory* **58**, 620 (2012).

[40] I. Kontoyiannis, Y. H. Lim, K. Papakonstantinopoulou, and W. Szpankowski, Compression and symmetry of small-world graphs and structures (2020), [arXiv:2007.15981](https://arxiv.org/abs/2007.15981).

[41] E. M. Wright, Graphs on unlabelled nodes with a large number of edges, *Proc. London Math. Soc.* **s3-28**, 577 (1974).

[42] T. Łuczak, How to deal with unlabeled random graphs, *J. Graph Theory* **15**, 303 (1991).

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

[44] N. Wormald, in *Proceedings of the International Congress of Mathematicians (ICM'18)*, Vol. 4 (World Scientific, Rio de Janeiro, 2019), pp. 3245–3264.

[45] A. E. Wegner and S. Olhede, Atomic subgraphs and the statistical mechanics of networks, *Phys. Rev. E* **103**, 042311 (2021).

[46] B. D. McKay and N. C. Wormald, Automorphisms of random graphs with specified vertices, *Combinatorica* **4**, 325 (1984).

[47] L. Brick, P. Gao, and A. Southwell, The threshold of symmetry in random graphs with specified degree sequences (2020), [arXiv:2004.01794](https://arxiv.org/abs/2004.01794).

[48] P. J. Cameron, The random graph, in *Math Paul Erdős II* (Springer, New York, 2013), pp. 353–378.

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

[50] B. Söderberg, General formalism for inhomogeneous random graphs, *Phys. Rev. E* **66**, 066121 (2002).

[51] B. Söderberg, Random graphs with hidden color, *Phys. Rev. E* **68**, 015102 (2003).

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

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

[54] O. Kallenberg, *Random Measures, Theory and Applications*, Probability Theory and Stochastic Modelling, Vol. 77 (Springer International Publishing, Cham, 2017).

[55] P. Hanlon, Counting interval graphs, *Trans. Amer. Math. Soc.* **272**, 383 (1982).

[56] R. P. Stanley and S. Fomin, *Enumerative Combinatorics*, Vol. 2 (Cambridge University Press, Cambridge, UK, 1999).

[57] G. Last and M. Penrose, *Lectures on the Poisson Process* (Cambridge University Press, Cambridge, UK, 2017).

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

[59] D. Krioukov and M. Ostilli, Duality between equilibrium and growing networks, *Phys. Rev. E* **88**, 022808 (2013).

[60] T. Łuczak, A. Magnier, and W. Szpankowski, Asymmetry and structural information in preferential attachment graphs, *Random Struct Alg.* **55**, 696 (2019).

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

[62] B. D. McKay and A. Piperno, Practical graph isomorphism, II, *J. Symb. Comput.* **60**, 94 (2014).

[63] N. C. Wormald, Generating random unlabelled graphs, *SIAM J. Comput.* **16**, 717 (1987).

[64] D. N. Hoover, *Relations on Probability Spaces and Arrays of Random Variables*, Technical Report, Institute for Advanced Study, Princeton, NJ (1979).

[65] D. J. Aldous, Representations for partially exchangeable arrays of random variables, *J. Multivar. Anal.* **11**, 581 (1981).