Letter

Construction of simplicial complexes with prescribed degree-size sequences

Tzu-Chi Yen*
Department of Computer Science, University of Colorado, Boulder, Colorado 80309, USA

(Received 31 May 2021; accepted 30 September 2021; published 28 October 2021)

We study the realizability of simplicial complexes with a given pair of integer sequences, representing the node degree distribution and the facet size distribution, respectively. While the s-uniform variant of the problem is NP-complete when $s \geqslant 3$, we identify two populations of input sequences, most of which can be solved in polynomial time using a recursive algorithm that we contribute. Combining with a sampler for the simplicial configuration model [J.-G. Young *et al.*, Phys. Rev. E **96**, 032312 (2017)], we facilitate the efficient sampling of simplicial ensembles from arbitrary degree and size distributions. We find that, contrary to expectations based on dyadic networks, increasing the nodes' degrees reduces the number of loops in simplicial complexes. Our work unveils a fundamental constraint on the degree-size sequences and sheds light on further analyses of higher-order phenomena based on local structures.

DOI: 10.1103/PhysRevE.104.L042303

Capturing higher-order dependencies is essential for the structural interpretation of the organization and behavior of complex systems [1–3]. Simplicial complex modeling, among other methods in applied topology [4–6], provides a combinatorial description of the topology of the system, featuring algebraic redundancies that may be compressed out using equivalence relations. Similar to networks, simplicial complexes are composed of nodes that represent system observables, and high-dimensional analogs of edges, called simplices, that represent polyadic relationships among the nodes. Simplicial modeling made several recent discoveries in complex systems, including the emergence of discontinuous transitions [7] in epidemic spreading [8,9] and synchronization [10,11], the multiscale hierarchy in adaptive voter models [12], the role of simplex size fluctuations in temporal networks [13], localization of dynamics [14], and percolation transitions [15–17]. In a related role, simplicial complexes have been used to summarize static features, addressing questions about the local geometry of data, such as in distinguishing the voting patterns in densely populated cities [18] and in describing the shape of scientific collaborations [19].

Dynamical processes on networks depend crucially on the network structure [20,21]. However, their generalizations to simplicial structures and nonpairwise interactions are relatively less understood. Notably, the basic question of which degree-size sequences can be realized by a complex is still open. In this Letter we make progress in this direction by extensive numerical experiments.

Let V be a finite set of nodes. A simplicial complex on V is a collection K of nonempty subsets of V, called simplices, subject to two requirements: First, for each node $v \in V$, the singleton $\{v\} \in K$; second, for all simplices $\tau \in K$, all its proper subsets $\sigma \subset \tau$ must also be in K. A *facet* is a maximal simplex, i.e., one that is not a subset of any other simplex. Note that a simplicial complex can be fully specified by listing only its facets, and we follow that convention here. We define the degree d_i of a node v_i as the number of facets incident on v_i and the size (cardinality) s_j of a facet σ_j as the number of nodes it contains. For any simplicial complex K, there is a corresponding degree-size sequence $\boldsymbol{t}_K = (\boldsymbol{d}, \boldsymbol{s})$, where $\boldsymbol{d} = \{d_1, d_2, \ldots, d_n\}$ and $\boldsymbol{s} = \{s_1, s_2, \ldots, s_m\}$ (see Fig. 1). However, the reverse statement is not always true. There are sequences $\boldsymbol{t}$ that cannot be realized by any simplicial complex. Hence, inspired by Ref. [22], we pose the *simplicial complex realization* problem: Given integer sequences $\boldsymbol{t} = (\boldsymbol{d}, \boldsymbol{s})$, does there exist a simplicial complex K with that degree-size sequence? When the answer is affirmative, we call the sequence simplicial.

The degree-size sequence reflects the local coupling patterns of the system. Models that constrain these features can often be used as null models that detect salient structures. In particular, the simplicial configuration model (SCM) [22] specifies a distribution of simplicial complexes with fixed degree-size sequences, and can be sampled via a Markov

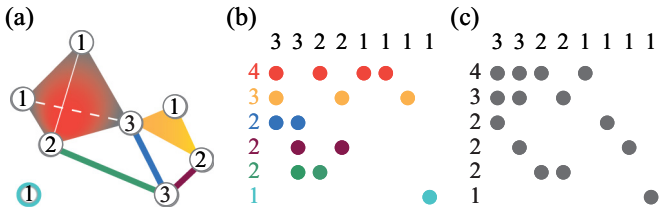

FIG. 1. (a) Geometric representation $|K|$ of a simplicial complex K with degree-size sequence $\boldsymbol{d} = \{3, 3, 2, 2, 1, 1, 1, 1\}$ and $\boldsymbol{s} = \{4, 3, 2, 2, 2, 1\}$. The numbers on the nodes represent their degrees. (b) Its incidence matrix, where the circles represent 1's and each row (respectively column) constitutes a facet (respectively node). (c) An alternative realization of the same sequence, following the algorithm described in the main text.

*tzuchi.yen@colorado.edu

chain Monte Carlo (MCMC) algorithm. The SCM extends the configuration model for graphs [23,24] and similar efforts in simplicial complexes of equal-size facets [25]. Critically, the MCMC requires an initial simplicial configuration to work, restricting its use to empirical data which can act as the initialization. The algorithm we propose yields an initialization for arbitrary simplicial sequences, not necessarily taken from an observed data set.

Related work. A key difficulty in *simplicial complex realization* is that no facet is allowed to be completely included in any other facet; we call this the "no-inclusion constraint." In contrast, hypergraphs have no such constraint. For simple hypergraphs, Ref. [26] gives a rejection sampling algorithm that samples realizations with given degrees and hyperedge sizes. For their nonsimple counterparts, Ref. [27] considers an MCMC approach for generating such hypergraphs and Ref. [28] ensures the existence of a starting realization which, however, needs not be simplicial. If we relax the notion of facets and consider instead the degree sequence of nodes partaking in given motifs, Refs. [29,30] use generating functions to study such networks.

Testing whether a degree-size sequence is simplicial is a generalization of the *graph realization* problem, in which the main result is the Erdős–Gallai theorem [31] that exactly characterizes graphical sequences with a set of easy-to-test inequalities. Equivalently, a particular graph realization can be constructed by a recursive application of the Havel–Hakimi theorem [32]. The reformulation of these theorems expedites the direct sampling of networks and facilitates understanding of their properties (e.g., all scale-free networks are sparse [33]). Moreover, networks enjoy tractable expressions for many ensemble-averaged quantities of interest (e.g., degree correlation, which tends to be disassortative in heterogeneous networks [34,35]), precisely because they are free from the no-inclusion constraint. For simpliciality testing, however, none of these methods applies. Recently, Deza, Levin, Meesum, and Onn [36] proved that deciding whether a given sequence is the degree sequence of a simple s-uniform hypergraph is NP-complete when $s \geqslant 3$, through a reduction from the *3-partition* problem [37]. Interestingly, simple s-uniform hypergraphs are equivalent to equidimensional simplicial complexes, because when all hyperedges are the same size, they automatically satisfy the no-inclusion constraint. This implies that deciding simpliciality *is* hard and there may not exist an efficient algorithm to enumerate and sample these instances. Yet, as our exploration reveals, not all sequence combinations are equally hard. For example, for any $\boldsymbol{s}$, taking the trivial degree sequence $\boldsymbol{d} = \{1, \ldots, 1\}$ immediately yields a simplicial realization.

In this Letter, we develop a deterministic, backtracking-based search algorithm that always correctly solves simpliciality, and present evidence that on most instances it runs in polynomial time. We then explore the topology of the constructed complex more generally via its Betti numbers, as well as using it as a seed for the SCM ensemble. With the randomized realizations, we reveal the regimes in which their expected topology changes as a function of the degree and size sequences.

Algorithm. Our algorithm proceeds as follows. It is given as input a node degree sequence $\boldsymbol{d}$ and a facet size sequence $\boldsymbol{s}$, where both sequences are nonincreasing. Let $n = |\boldsymbol{d}|$ and $m = |\boldsymbol{s}|$, where $|\cdot|$ stands for the cardinality. Simpliciality fails trivially if there are fewer 1's in $\boldsymbol{d}$ than in $\boldsymbol{s}$, as it is doomed to violate the no-inclusion constraint, or if $d_1 > m$ or $s_1 > n$, or $\sum \boldsymbol{d} \neq \sum \boldsymbol{s}$, as it would be impossible to form an incidence matrix. As a preprocessing step, we pair the 1's in $\boldsymbol{d}$ with those in $\boldsymbol{s}$ to make a partial output.

Next, we pick s_1 nodes to make a candidate facet $\hat{\sigma}^1$. This selection is not done stochastically, but favors higher-degree nodes—the candidate facet $\hat{\sigma}$ with the largest sum of input degrees $w(\hat{\sigma}) = \sum_{i \in \hat{\sigma}} d_i$ is preferred. We ensure that $\hat{\sigma}$ is not included in any existing facet, or we proceed with the next one in heuristic order. We will validate $\hat{\sigma}^1$ with a series of rules. If it fails any of them, we take the next candidate, until we accept and advance to pick s_2 nodes for $\hat{\sigma}^2$. For $\hat{\sigma}^\ell$ ($\ell \geqslant 2$), the facets with larger $w(\hat{\sigma}^\ell)$ still attain precedence.

With this overall structure in mind, we now express the algorithm recursively. At each branching stage ℓ, the input is a 3-tuple $(\boldsymbol{d}^\ell, \boldsymbol{s}^\ell, \boldsymbol{B}^\ell)$, where $\boldsymbol{d}^\ell$ is the residual degree sequence and $\boldsymbol{s}^\ell$ the residual size sequence, denoting the marginal sums of the incidence matrix that still need to be fulfilled. In addition, we have a collection of "blocking" facets $\boldsymbol{B}^\ell := \{\sigma^k : 1 \leqslant k < \ell\}$, where each accepted facet σ^k plays a role in the no-inclusion constraint. After accepting $\hat{\sigma}^\ell$, the output is again a 3-tuple $(\boldsymbol{d}^{\ell+1}, \boldsymbol{s}^{\ell+1}, \boldsymbol{B}^{\ell+1})$. The algorithm returns a simplicial realization if $\sum \boldsymbol{s}^\ell = 0$, or a negative result when the entire search tree has been traversed.

To validate a candidate facet $\hat{\sigma}^\ell$, we assume that $\hat{\sigma}^\ell$ is accepted and virtually move to the next branching stage to obtain a number of intermediate data, which must obey the validation rules before we actually branch. Precisely, we compute the 3-tuple $(\boldsymbol{d}^{\ell+1}, \boldsymbol{s}^{\ell+1}, \boldsymbol{B}^{\ell+1})$ at stage $\ell + 1$ and the set of nonshielding nodes $Q^\ell := V^\ell \setminus \hat{\sigma}^\ell$, where V^ℓ is the set of nodes at stage ℓ. For clarity, we drop all superscripts in the following developments.

Rule 1 (for $\boldsymbol{d}$ and $\boldsymbol{s}$) [38]. We ask that $d_{\max} \leqslant |\boldsymbol{s}|$, and that $s_{\max} \leqslant |\boldsymbol{d}|$, where equality in the latter requirement is only allowed when $|\boldsymbol{s}| = 1$.

Rule 2 (for $\boldsymbol{d}$, $\boldsymbol{s}$, and Q). We require that $|\boldsymbol{s}| \leqslant \sum_{i \in Q} d_i$, as every subsequent facet must contain at least one nonshielding node (i.e., element of Q) in order to not be included in $\hat{\sigma}$.

If equality holds, each subsequent facet must take exactly one nonshielding node. To secure its availability, we can thus further require that $s_{\max} - 1 \leqslant |\boldsymbol{d}| - |Q|$.

Rule 3 (for $\boldsymbol{d}$ and $\boldsymbol{B}$). Let V' be the set of nodes with a positive residual degree. We require that, for every blocking facet $\sigma \in \boldsymbol{B}$, V' be not a subset of σ.

While rule 1 is essentially the precondition, rules 2 and 3 are meant to cut the combinatorial explosion as much as possible, but not prohibit any realizable sequence from being realized. If accepting a candidate facet leads to $d_i = |\boldsymbol{s}|$ for any node i, these nodes are required to be used for the remaining facets—we invoke a subroutine to recursively consume those "forced" degrees. A Python implementation of the algorithm is freely available [39]. The pseudocode is given in the Supplemental Material [40].

Easy and hard instances. To benchmark the algorithm, we define the running time as

$$\tau = \tau_{\mathrm{b}} + \tau_{\mathrm{r}} ,$$

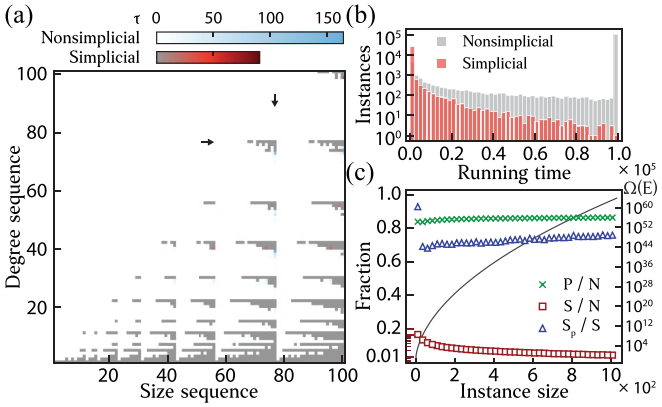

FIG. 2. (a) Realizability of all degree-size sequences with partitions of 13, following ascending compositions. Color bars indicate running time τ, where white to bluish colors mark the nonsimplicial instances and gray to reddish colors mark the simplicial ones. Around 17% of the instances are hard. Qualitatively, more uniform degree sequences can pair with more size sequences to be simplicial and vice versa—see the arrows, which correspond to a sequence of $\{3, 2, 2, 2, 2, 2\}$. (b) Running time distribution when size sequence is fixed at $\boldsymbol{s} = \{3, 3, \ldots, 3\}$ with $|\boldsymbol{s}| = 20$, with $\boldsymbol{d}$ set to all partitions of 60. Around 84% of the instances are easy (not shown). (c) Simplicial (□) and polynomial (×; among all simplicial: △) fractions vs instance size (bottom part in log scale) for $N = 10^6$ uniform random partitions [41,42]. The solid line shows the number of potential inputs at a specific E, i.e., $\Omega(E) = a(E) \times a(E)$, where $a(n)$ is the number of partitions of n. In all cases, we apply a cutoff at $\tau = 10^5$.

where τ_b records the number of times the algorithm backtracks (due to candidate depletion) and τ_r records the number of times a proposed facet is rejected. These numbers are correlated: If we come up with better validation rules, we reduce rejections and prevent backtracking. We call a degree-size sequence easy (or polynomial) if $\tau = 0$, meaning that no backtracking is necessary, the algorithm either finds a realization in near-linear time or rejects simpliciality immediately. Otherwise, we call it hard. Of course, this distinction is dependent on the choice of algorithm and "hard" instances with τ small will still be easy in practice, but for the purposes of understanding the problem, we find it useful to distinguish between those cases that are solved greedily by our algorithm and those that are not.

In Fig. 2(a), we show the realizability of all combinations of degree-size sequences of a fixed instance size $E = \sum \boldsymbol{d} = 13$. The self-similar pattern reflects the recursive nature of the algorithm. However, we are unable to conclude a recurrence relation in the spirit of Erdős–Gallai, due to the inherent complexity of solving particular instances. Indeed, there are two distinct populations of easy and hard instances, where a major fraction of inputs falls in the polynomial region. The easy majority can be understood as the iterative application of the heuristic policy. For 3-uniform, NP-complete instances [Fig. 2(b)], most inputs are polynomial and, on the average, the nonsimplicial instances are harder than simplicial ones, as tree exhaustion is needed. This highlights a useful property that false negatives are kept at a minimum when applying a reasonable cutoff.

To investigate the dependence of the simpliciality and hardness of degree-size sequences on the instance size, we perform extensive numerical calculations, generating uniform ensembles of sequences of random integers, $(\boldsymbol{d}, \boldsymbol{s})$, with a range up to $E = 1010$. We test each sequence for simpliciality by applying the algorithm and compute for each E the simplicial fraction S/N, where S is the total number of simplicial sequences in the ensemble and N is the ensemble size, chosen at $N = 10^6$. We also compute the fractions P/N and S_p/S, where P is the total number of polynomial instances in the ensemble and S_p is the number of instances that are both simplicial and polynomial. The results, plotted in Fig. 2(c), clearly demonstrate the persistent existence of easy and hard populations at much larger sizes. An open question is to what extent we can further separate these classes in polynomial time, perhaps through better validation rules.

Heuristic policy and topology. When data are encoded as a simplicial complex K, we can characterize their homotopical information by the Betti numbers $\beta_k(K)$ [5,43]. They quantify the k-dimensional connectivity of objects by comparing their volume and boundary at each dimension—β_0 is the number of connected components, β_1 the number of homological cycles (or loops), while higher β_k effectively counts the number of k-dimensional cavities. The topological cavities can be geometric in physical space, such as in granular packings [44], or abstract structures in experimental measurements [18,19]. For instance, a cavity in diffusion magnetic resonance imaging readings could indicate axonal dropout, a neurological disorder [45]. In the following, we focus on the first two Betti numbers because they are easier to interpret.

An important feature of our heuristic policy is that high-degree nodes tend to form larger facets, resulting in a core-periphery structure [46] with dangling and isolated facets on the fringe. Therefore, the heuristic tends to find a realization with a relatively large number of connected components and few loops [47]. We show in Fig. 3(a) this feature on an empirical degree-size sequence extracted from the human diseasome network [22,48]. Indeed, our algorithm can discover a realization with $\beta_0 = 643$ and $\beta_1 = 5$ compatible with the degree-size sequence, whereas typical realizations sampled

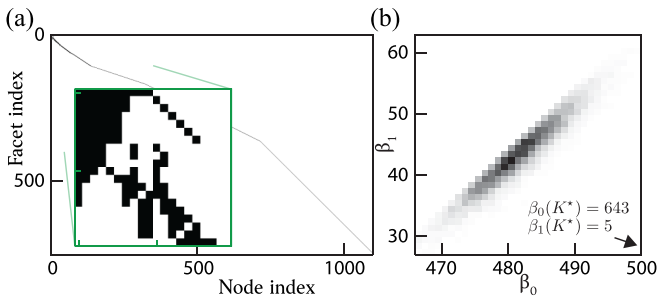

FIG. 3. (a) Realization $K^\star$ of the degree-size sequence from the human diseasome network (after pruning included faces) [22], showing an assortative degree structure. Black squares show which nodes make which facet, with $n = 1100$ and $m = 752$. The indices are sorted such that nodes (respectively facets) with a larger degree (respectively size) have a lower index. The inset zooms in on the composition of the largest 20 facets. (b) Joint Betti number distribution of 10^4 randomized realizations of $K^\star$.

from the SCM have much lower β_0 and much higher β_1 [see Fig. 3(b)]. This algorithmic trait is consistent across other data sets we examined [39]. More broadly, this priority policy shows a minimal example where adding degree-size correlations can introduce atypical Betti numbers. This sheds light on growth mechanisms that generate structures with specific homology, such as being totally connected [49] or containing many cycles [50].

SCM ensembles. To test the method, we generate random degree-size sequences and test them for simpliciality. We generate from two Poisson distributions, modified so that all facets of size 1 are guaranteed to be matched with a degree-1 node. For each simplicial sequence, we use the constructed complex to initialize an MCMC sampler for the SCM ensemble [22] and compute its mean Betti numbers $\bar{\boldsymbol{\beta}}_0$ and $\bar{\boldsymbol{\beta}}_1$. Then we take the average over the random partitions. Note that finding an initialization is inefficient with a rejection sampler. The result, shown in Fig. 4, is a systematic study of $\langle\bar{\boldsymbol{\beta}}_0\rangle$ and $\langle\bar{\boldsymbol{\beta}}_1\rangle$ with respect to $\bar{\boldsymbol{d}}$ and $\bar{\boldsymbol{s}}$, where $\bar{\boldsymbol{s}}$ is the mean facet size and $\bar{\boldsymbol{d}}$ is the mean node degree, excluding the nodes that are created to match the facets.

In Fig. 4(a), we observe that the average number of connected components $\langle\bar{\boldsymbol{\beta}}_0\rangle$, in the SCM ensemble, decreases with increasing $\bar{\boldsymbol{s}}$, likely due to the reduction of facets with cardinality 1 in the size sequence. However, the distribution of cycles $\langle\bar{\boldsymbol{\beta}}_1\rangle$ is considerably more complicated. In the low $\bar{\boldsymbol{s}}$ regime, the simplicial complex acts as a dyadic network, where denser networks contain more loops. By contrast, in the high $\bar{\boldsymbol{s}}$ regime, the system is abundant in large simplices. Once there is a fraction of higher-degree nodes, we have no other option but to bundle the large facets with those nodes, resulting in a treelike, few loops complex. We supply a parallel analysis on d-regular degree distributions in Fig. 4(b), which tend to entail more loops than Poissonian ones with the same average degree, as they possess fewer high-degree nodes.

The decay of the average number of cycles $\langle\bar{\boldsymbol{\beta}}_1\rangle$ when the average degree $\bar{\boldsymbol{d}}$ is increased is reminiscent of the law of large numbers for Betti numbers in random simplicial complexes (e.g., the Linial–Meshulam model [51] or the random clique complex [52]). We note that in these studies, the limiting behavior of Betti numbers is discussed in the context of increasing facet density, where the decay is driven by filling the k-dimensional cycles with $(k+1)$-simplices. Here, the system has a fixed number of simplices and the decay is driven by the no-inclusion constraints that prevent the realization of any such complexes.

We also observe that $\langle\bar{\boldsymbol{\beta}}_1\rangle$ is unimodal with respect to mean facet size $\bar{\boldsymbol{s}}$ [Figs. 4(a) and 4(b)]. The unimodality comes from the competitive relationship between the gradually multifaceted local geometry and the depletion of available facets. When $\bar{\boldsymbol{s}}$ is small, there are fewer large facets to avoid for smaller ones and, thus, increasing the facet sizes has a similar effect as densifying the degrees, which creates more loops [cf. Fig. 4(c), ① to ②]. That said, the loops are destroyed as facets merge into larger ones [cf. Fig. 4(c), ② to ③]. This suggests the existence of an optimal facet size distribution for loopy configurations, as the number of facets is limited.

Discussion. In computational complexity, many graph problems are NP-hard in general, but may be in P for certain classes of graphs [53]. *Simplicial complex realization* is yet another addition to the list. We present a depth-bounded branching algorithm whose complexity presents a rich structure. In particular, simpliciality can be solved in time $f(m)E^c$ for some constant $c \approx 1$, where E^c is the time spent in validations and the prefactor counts the number of nodes in the execution tree. For easy instances, $f(m) = m$, and the algorithm is linear in instance size. Otherwise, we have $f(m) = b^{m+1}$ in general, where b is the branching ratio that grows with instance size. We find that b is highly heterogeneous as a function of the branching stage—the searcher stalls at midstages, not at the beginning or the end. It remains open to accurately parametrize the complexity of hard instances of the simpliciality problem, and to prove rigorously, if the algorithm runs in linear time on average. Finally, we note that the branching design has opened up an avenue to systematically improve the algorithm, for example, through stronger validation rules to reduce the branching ratio, or introducing variants by nonchronological backjumping or clause learning techniques, as critically used in modern Boolean satisfiability (SAT) solvers [54].

Aside from these computational complexity questions, the boundary between easy and hard instances deserves further attention. We find that the instances tend to be harder when $(\boldsymbol{d}, \boldsymbol{s})$ contain numerous uniform entries, whereas a Poisson-Poisson combination yields very little backtracking.

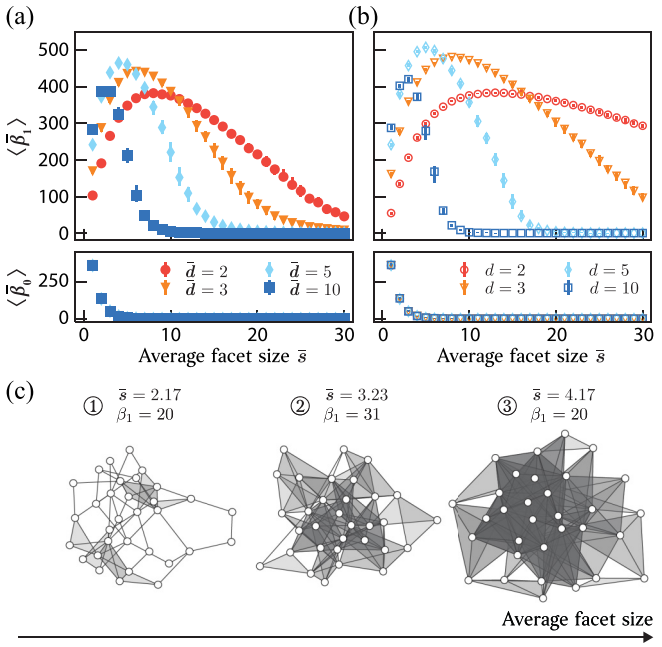

FIG. 4. (a) Average of the first two Betti numbers $\langle\bar{\boldsymbol{\beta}}_0\rangle$ and $\langle\bar{\boldsymbol{\beta}}_1\rangle$ of the simplicial complexes with Poisson-Poisson degree-size sequences, (b) Poisson size sequence and d-regular degree sequence. Each point shows the median of 10^2 replicates of the indicated ensemble (see legend) and error bars show 25%–75% quantiles. For each realization in the ensemble, we compute the average of Betti numbers from 10 SCM replicates. All complexes have $E = 10^3$. (c) Sketches of the simplicial structure. Enlarging the facets, while fixing the degree sequence, will first increase [① to ②] and then decrease [② to ③] the number of loops. The complexes have the same degree distribution at $\bar{\boldsymbol{d}} = 2.86$.

The understanding of when and why their hardness differs is poor compared to what is known for constraint satisfaction problems [55] or the inference of stochastic block models [56]. This raises a number of open questions, for example, is there an algorithmic phase transition that separates easy and hard regions? Here, hard instances could mean either $\tau > 0$ or *really hard* in some other sense, e.g., takes exponential time, as seen in SAT. Or, would there even be two different phase transitions? It is also known that *graph isomorphism* can be solved in linear time for random graphs [57,58], by leveraging the fact that in random graphs the degree distribution is essentially never uniform, so that high-degree nodes help break symmetries. For simpliciality, it could be useful to investigate the dependency of algorithmic hardness on the degree sequence among equidimensional sequences. We believe that the solutions to these problems will require new insights in the statistical physics of computation [59,60], notably, to identify the "order parameter" that characterizes the algorithmic barrier to hard instances [61].

In closing, we have developed a constructive algorithm to allow faster realization of simplicial complexes with arbitrary degree-size sequences. Our algorithm paves the way for a more comprehensive analysis of higher-order phenomena in terms of local structural attributes, revealing their roles in various dynamical systems.

Acknowledgments. I am particularly grateful to Joshua A. Grochow for significant discussions. I am grateful to Alice Patania and Jean-Gabriel Young for helpful correspondence, as well as to Daniel B. Larremore for support. I acknowledge the BioFrontiers Computing Core at the University of Colorado Boulder for computing facilities. This work was funded in part by Grochow startup funds.

[1] L. Torres, A. S. Blevins, D. Bassett, and T. Eliassi-Rad, The why, how, and when of representations for complex systems, SIAM Rev. **63**, 435 (2021).
[2] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, Networks beyond pairwise interactions: Structure and dynamics, Phys. Rep. **874**, 1 (2020).
[3] C. Bick, E. Gross, H. A. Harrington, and M. T. Schaub, What are higher-order networks? arXiv:2104.11329.
[4] H. Edelsbrunner and J. L. Harer, *Computational Topology: An Introduction* (American Mathematical Society, Providence, RI, 2010).
[5] R. Ghrist, *Elementary Applied Topology*, 1st ed. (CreateSpace, Scotts Valley, CA, 2014).
[6] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, A roadmap for the computation of persistent homology, EPJ Data Sci. **6**, 17 (2017).
[7] C. Kuehn and C. Bick, A universal route to explosive phenomena, Sci. Adv. **7**, eabe3824 (2021).
[8] I. Iacopini, G. Petri, A. Barrat, and V. Latora, Simplicial models of social contagion, Nat. Commun. **10**, 2485 (2019).
[9] N. W. Landry and J. G. Restrepo, The effect of heterogeneity on hypergraph contagion models, Chaos **30**, 103117 (2020).
[10] P. S. Skardal and A. Arenas, Higher order interactions in complex networks of phase oscillators promote abrupt synchronization switching, Commun. Phys. **3**, 218 (2020).
[11] A. P. Millán, J. J. Torres, and G. Bianconi, Explosive Higher-Order Kuramoto Dynamics on Simplicial Complexes, Phys. Rev. Lett. **124**, 218301 (2020).
[12] L. Horstmeyer and C. Kuehn, Adaptive voter model on simplicial complexes, Phys. Rev. E **101**, 022305 (2020).
[13] G. Petri and A. Barrat, Simplicial Activity Driven Model, Phys. Rev. Lett. **121**, 228301 (2018).
[14] G. St-Onge, V. Thibeault, A. Allard, L. J. Dubé, and L. Hébert-Dufresne, Social Confinement and Mesoscopic Localization of Epidemics on Networks, Phys. Rev. Lett. **126**, 098301 (2021).
[15] G. Bianconi and R. M. Ziff, Topological percolation on hyperbolic simplicial complexes, Phys. Rev. E **98**, 052308 (2018).
[16] F. A. N. Santos, E. P. Raposo, M. D. Coutinho-Filho, M. Copelli, C. J. Stam, and L. Douw, Topological phase transitions in functional brain networks, Phys. Rev. E **100**, 032414 (2019).
[17] O. Bobrowski and P. Skraba, Homological percolation and the Euler characteristic, Phys. Rev. E **101**, 032304 (2020).
[18] M. Feng and M. A. Porter, Persistent homology of geospatial data: A case study with voting, SIAM Rev. **63**, 67 (2021).
[19] A. Patania, G. Petri, and F. Vaccarino, The shape of collaborations, EPJ Data Sci. **6**, 18 (2017).
[20] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, UK, 2008).
[21] M. Porter and J. Gleeson, *Dynamical Systems on Networks: A Tutorial* (Springer, Cham, 2016).
[22] J.-G. Young, G. Petri, F. Vaccarino, and A. Patania, Construction of and efficient sampling from the simplicial configuration model, Phys. Rev. E **96**, 032312 (2017).
[23] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, Eur. J. Comb. **1**, 311 (1980).
[24] B. K. Fosdick, D. B. Larremore, J. Nishimura, and J. Ugander, Configuring random graph models with fixed degree sequences, SIAM Rev. **60**, 315 (2018).
[25] O. T. Courtney and G. Bianconi, Generalized network structures: The configuration model and the canonical ensemble of simplicial complexes, Phys. Rev. E **93**, 062311 (2016).
[26] M. Dyer, C. Greenhill, P. Kleer, J. Ross, and L. Stougie, Sampling hypergraphs with given degrees, arXiv:2006.12021.
[27] P. S. Chodrow, Configuration models of random hypergraphs, J. Complex Netw. **8**, cnaa018 (2020).
[28] N. A. Arafat, D. Basu, L. Decreusefond, and S. Bressan, Construction and random generation of hypergraphs with prescribed degree and dimension sequences, arXiv:2004.05429.
[29] B. Karrer and M. E. J. Newman, Random graphs containing arbitrary distributions of subgraphs, Phys. Rev. E **82**, 066118 (2010).
[30] P. Mann, V. A. Smith, J. B. O. Mitchell, and S. Dobson, Random graphs with arbitrary clustering and their applications, Phys. Rev. E **103**, 012309 (2021).
[31] P. Erdős and T. Gallai, Graphs with prescribed degrees of vertices (Hungarian), Mat. Lopak **11**, 264 (1960).

[32] V. Havel, A remark on the existence of finite graphs (Czech), Časopis Pěst. Mat. **80**, 477 (1955); S. L. Hakimi, On realizability of a set of integers as degrees of the vertices of a linear graph. I, J. Soc. Ind. Appl. Math. **10**, 496 (1962).

[33] C. I. Del Genio, T. Gross, and K. E. Bassler, All Scale-Free Networks are Sparse, Phys. Rev. Lett. **107**, 178701 (2011).

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

[35] S. Johnson, J. J. Torres, J. Marro, and M. A. Muñoz, Entropic Origin of Disassortativity in Complex Networks, Phys. Rev. Lett. **104**, 108702 (2010).

[36] A. Deza, A. Levin, S. M. Meesum, and S. Onn, Optimization over degree sequences, SIAM J. Disc. Math. **32**, 2067 (2018).

[37] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (W. H. Freeman, New York, 1979).

[38] This rule is similar to, but stronger than, the Gale–Ryser characterization of bigraphic pairs of sequences [62], which would allow $s_{\max} = |\boldsymbol{d}|$ for nonterminal facets, thus violating the no-inclusion constraint.

[39] T.-C. Yen, The `simplicial-test` Python library, version 1.3, https://pypi.org/project/simplicial-test.

[40] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.104.L042303 for a detailed description of the algorithm.

[41] A. Nijenhuis and H. S. Wilf, *Combinatorial Algorithms for Computers and Calculators*, 2nd ed., Computer Science and Applied Mathematics (Academic, New York, 1978).

[42] The Sage Developers, SageMath, the Sage Mathematics Software System, 2021, version 9.2, https://www.sagemath.org.

[43] A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, UK, 2002).

[44] M. Saadatfar, H. Takeuchi, V. Robins, N. Francois, and Y. Hiraoka, Pore configuration landscape of granular crystallization, Nat. Commun. **8**, 15082 (2017).

[45] A. E. Sizemore, J. E. Phillips-Cremins, R. Ghrist, and D. S. Bassett, The importance of the whole: Topological data analysis for the network neuroscientist, Netw. Neurosci. **3**, 656 (2018).

[46] R. J. Gallagher, J.-G. Young, and B. F. Welles, A clarified typology of core-periphery structure in networks, Sci. Adv. **7**, eabc9800 (2021).

[47] The latter case is due to an algorithmic choice to favor the facets with small residual degrees, from among the candidates with the same priority w [40].

[48] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabási, The human disease network, Proc. Natl. Acad. Sci. U.S.A. **104**, 8685 (2007).

[49] S. Horvat and C. D. Modes, Connectedness matters: Construction and exact random sampling of connected networks, J. Phys.: Complex. **2**, 015008 (2020).

[50] M. Chujyo and Y. Hayashi, A loop enhancement strategy for network robustness, Appl. Netw. Sci. **6**, 3 (2021).

[51] N. Linial and Y. Peled, On the phase transition in random simplicial complexes, Ann. Math. **184**, 745 (2016).

[52] S. Kanazawa, Law of large numbers for Betti numbers of homogeneous and spatially independent random simplicial complexes, Random Struct. Algorithms (2021), doi:10.1002/rsa.21015.

[53] V. E. Alekseev, On easy and hard hereditary classes of graphs with respect to the independent set problem, Discrete Appl. Math. **132**, 17 (2003).

[54] J. Marques-Silva, I. Lynce, and S. Malik, Conflict-driven clause learning SAT solvers, in *Handbook of Satisfiability* (IOS Press, Amsterdam, 2009), pp. 131–153.

[55] F. Krzakala and L. Zdeborová, Hiding Quiet Solutions in Random Constraint Satisfaction Problems, Phys. Rev. Lett. **102**, 238701 (2009).

[56] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications, Phys. Rev. E **84**, 066106 (2011).

[57] L. Babai and L. Kucera, Canonical labelling of graphs in linear average time, in *20th Annual Symposium on Foundations of Computer Science (sfcs 1979)* (IEEE, New York, 1979), pp. 39–46.

[58] L. Babai, P. Erdős, and S. M. Selkow, Random graph isomorphism, SIAM J. Comput. **9**, 628 (1980).

[59] L. Zdeborová and F. Krzakala, Statistical physics of inference: Thresholds and algorithms, Adv. Phys. **65**, 453 (2016).

[60] C. Moore and S. Mertens, *The Nature of Computation*, 1st ed. (Oxford University Press, Oxford, UK, 2011).

[61] P. Cheeseman, B. Kanefsky, and W. M. Taylor, Where the really hard problems are, in *Proceedings of the 12th International Joint Conference on Artificial Intelligence - Volume 1*, IJCAI'91 (Morgan Kaufmann Publishers, San Francisco, CA, 1991), pp. 331–337.

[62] D. Gale, A theorem on flows in networks, Pac. J. Math. **7**, 1073 (1957); H. J. Ryser, Combinatorial properties of matrices of zeros and ones, Can. J. Math. **9**, 371 (1957).