Symmetric Graph Properties Have Independent Edges

Dimitris Achlioptas[1] ⋆ and Paris Siminelakis[2] ⋆⋆

[1] Department of Computer Science
University of California, Santa Cruz
optas@cs.ucsc.edu
[2] Department of Electrical Engineering
Stanford University
psimin@stanford.edu

Abstract. In the study of random structures we often face a trade-off between realism and tractability, the latter typically enabled by independence assumptions. In this work we initiate an effort to bridge this gap by developing tools that allow us to work with independence without assuming it. Let $\mathcal{G}_n$ be the set of all graphs on n vertices and let S be an arbitrary subset of $\mathcal{G}_n$, e.g., the set of all graphs with m edges. The study of random networks can be seen as the study of properties that are true for *most* elements of S, i.e., that are true with high probability for a uniformly random element of S. With this in mind, we pursue the following question: *What are general sufficient conditions for the uniform measure on a set of graphs $S \subseteq \mathcal{G}_n$ to be well-approximable by a product measure on the set of all possible edges?*

1 Introduction

Since their introduction in 1959 by Erdős and Rényi [6] and Gilbert [8], respectively, $G(n, m)$ and $G(n, p)$ random graphs have dominated the mathematical study of random networks [2, 10]. Given n vertices, $G(n, m)$ selects uniformly among all graphs with m edges, whereas $G(n, p)$ includes each edge independently with probability p. A refinement of $G(n, m)$ are graphs chosen uniformly among all graphs with a given degree sequence, a distribution made tractable by the configuration model of Bollobás [2]. Due to their mathematical tractability these three models have become a cornerstone of Probabilistic Combinatorics and have found application in the Analysis of Algorithms, Coding Theory, Economics, Game Theory, and Statistical Physics.

⋆ Research supported by ERC Starting Grant StG-210743 and an Alfred P. Sloan Fellowship. Research co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: ARISTEIA II.
⋆⋆ Supported in part by an Onassis Foundation Scholarship.

At the foundation of this mathematical tractability lies symmetry: the probability of all edge sets of a given size is either the same, as in $G(n, p)$ and $G(n, m)$, or merely a function of the potency of the vertices involved, as in the configuration model. This extreme symmetry bestows numerous otherworldly properties, including near-optimal expansion. Perhaps most importantly, it amounts to a complete lack of geometry, as manifest by the fact that the shortest path metric of such graphs suffers maximal distortion when embedded in Euclidean space [14]. In contrast, vertices of real networks are typically embedded in some low-dimensional geometry, either explicit (physical networks), or implicit (social and other latent semantics networks), with distance being a strong factor in determining the probability of edge formation.

While the shortcomings of the classical models have long been recognized, proposing more realistic models is not an easy task. The difficulty lies in achieving a balance between realism and mathematical tractability: it is only too easy to create network models that are both ad hoc and intractable. By now there are thousands of papers proposing different ways to generate graphs with desirable properties [9] the vast majority of which only provide heuristic arguments to support their claims. For a gentle introduction the reader is referred to the book of Newman [17] and for a more mathematical treatment to the books of Chung and Lu [4] and of Durrett [5].

In trying to replicate real networks one approach is to keep adding features, creating increasingly complicated models, in the hope of *matching* observed properties. Ultimately, though, the purpose of any good model is prediction. In that sense, the reason to study (random) graphs with certain properties is to understand what *other* graph properties are (typically) implied by the assumed properties. For instance, the reason we study the uniform measure on graphs with m edges, i.e., $G(n, m)$, is to understand "what properties are typically implied by the property of having m edges" (and we cast the answer as "properties that hold with high probability in a 'random' graph with m edges"). Notably, analyzing the uniform measure even for this simplest property is non-trivial. The reason is that it entails the single massive choice of an m-subset of edges, rather than m independent choices. In contrast, the independence of choices in $G(n, p)$ makes that distribution far more accessible, dramatically enabling analysis.

Connecting $G(n, m)$ and $G(n, p)$ is a classic result of random graph theory. The key observation is that to sample according to $G(n, p)$, since edges are independent and equally likely, we can first sample an integer $m \sim \mathrm{Bin}\left(\binom{n}{2}, p\right)$ and then sample a uniformly random graph with m edges, i.e., $G(n, m)$. Thus, for $p = p(m) = m/\binom{n}{2}$, the random graph $G \sim G(n, m)$ and the two random graphs $G^{\pm} \sim G(n, (1 \pm \epsilon)p)$ can be coupled so that, viewing each graph as a set of edges, with high probability,

$$G^{-} \subseteq G \subseteq G^{+} \ . \tag{1}$$

The significance of this relationship between what we *wish* to study (uniform measure) and what we *can* study (product measure) can not be overestimated. It manifests most dramatically in the study of monotone properties: to study a monotone, say, increasing property in $G \sim G(n, m)$ it suffices to bound from above its probability in G^+ and from below in G^-. This connection has been thoroughly exploited to establish threshold functions for a host of monotone graph properties such as Connectivity, Hamiltonicity, and Subgraph Existence, making it the workhorse of random graph theory.

In this work we seek to extend the above relationship between the uniform measure and product measures to properties more delicate than having a given number of edges. In doing so we (i) provide a tool that can be used to revisit a number of questions in random graph theory from a more realistic angle and (ii) lay the foundation for designing random graph models eschewing independence assumptions. For example, our tool makes short work of the following set of questions (which germinated our work):

Given an arbitrary collection of n points on the plane what can be said about the set of all graphs that can be built on them using a given amount of wire, i.e., when connecting two points consumes wire equal to their distance? What does a uniformly random such graph look like? How does it change as a function of the available wire?

1.1 Our Contribution

A product measure on the set of all undirected simple graphs on n vertices, $\mathcal{G}_n$, is specified by a symmetric matrix $\mathbf{Q} \in [0,1]^{n \times n}$ where $Q_{ii} = 0$ for $i \in [n]$. By analogy to $G(n, p)$ we denote by $G(n, \mathbf{Q})$ the measure in which every edge $\{i, j\}$ is included independently with probability $Q_{ij} = Q_{ji}$. Let $S \subseteq \mathcal{G}_n$ be arbitrary. Our main result is a sufficient condition for the uniform measure on S, denoted by $U(S)$, to be approximable by a product measure in the following sense.

Sandwichability. *The measure $U(S)$ is (ϵ, δ)-sandwichable if there exists an $n \times n$ symmetric matrix $\mathbf{Q}$ such that the distributions $G \sim U(S)$ and $G^{\pm} \sim G(n, (1 \pm \epsilon)\mathbf{Q})$ can be coupled so that* $\Pr[G^- \subseteq G \subseteq G^+] \geq 1 - \delta$.

Informally, the two conditions required for our theorem to hold are as follows:

Partition Symmetry. The set S should be symmetric with respect to *some* partition $\mathcal{P} = (P_1, \ldots, P_k)$ of the $\binom{n}{2}$ possible edges. More specifically, for a partition $\mathcal{P}$ define the *edge profile* of a graph G with respect to $\mathcal{P}$ to be the k-dimensional vector $\mathbf{m}(G) = (m_1(G), \ldots, m_k(G))$ where $m_i(G)$ counts the number of edges in G from part P_i. Partition symmetry amounts to the requirement that the characteristic function of S can depend on *how many* edges are included from each part but not on *which* edges. That is, if we let $\mathbf{m}(S) := \{\mathbf{m}(G) : G \in S\}$, then $\forall G \in \mathcal{G}_n, \mathbb{I}_S(G) = \mathbb{I}_{\mathbf{m}(S)}(\mathbf{m}(G))$. The $G(n, m)$ model is recovered by considering the trivial partition with $k = 1$ parts and $\mathbf{m}(S) = \{m\}$. Far more interestingly, in our motivating example edges are partitioned into equivalence classes according to their cost $\mathbf{c}$ (distance of endpoints) and the characteristic function allows graphs whose edge profile $\mathbf{m}(G)$ does not violate the total wire budget $C_B = \{\mathbf{v} \in \mathbb{N}^k : \mathbf{c}^\intercal \mathbf{v} \leq B\}$. We discuss the motivation for edge-partition symmetry at length in Section 2.

Convexity. Since membership in S depends solely on a graph's edge-profile, it follows that a uniformly random element of S can be selected as follows: (i) select an edge profile $\mathbf{v} = (v_1, \ldots, v_k) \in \mathbb{R}^k$ from the distribution on $\mathbf{m}(S)$ induced by $U(S)$, and then (ii) for each $i \in [k]$ independently select a uniformly random v_i-subset of P_i. In other words, the complexity of the uniform measure on S manifests entirely in the *induced* distribution on $\mathbf{m}(S) \in \mathbb{N}^k$ whose structure we need to capture.

Without any assumptions the set $\mathbf{m}(S)$ can be arbitrary, e.g., S can be the set of graphs having either $n^{1/2}$ or $n^{3/2}$ edges, rendering any approximation by a product measure hopeless. To impose some regularity we require the discrete set $\mathbf{m}(S)$ to be *convex* in the sense, that *it equals the set of integral points in its convex hull*. While convexity is not strictly necessary for our proof method to work (see Section 5), we feel that it provides a clean conceptual framework while still allowing very general properties to be expressed. These include all properties expressible as Linear Programs in the number of edges from each part, but also properties involving non-linear constraints, e.g., the absence of percolation. (Our original example, of course, amounts to a single linear inequality constraint.) Most importantly, since convex sets are closed under intersection, convex properties can be composed arbitrarily while remaining amenable to approximability by a product measure.

We state our results formally in Section 4. The general idea is this.

Theorem 1 (Informal). *If S is a convex symmetric set, then $U(S)$ is sandwichable by a product measure $G(n, \mathbf{Q}^*)$.*

The theorem is derived by following the *Principle of Maximum Entropy*, i.e., by proving that the induced measure on the set of edge-profiles $\mathbf{m}(S)$ concentrates around a unique vector $\mathbf{m}^*$, obtained by solving an entropy (concave function) maximization problem on the convex hull of $\mathbf{m}(S)$. The maximizer $\mathbf{m}^*$ can in many cases be computed explicitly, either analytically or numerically, and the product measure $\mathbf{Q}^*$ follows readily from it. Indeed, the maximizer $\mathbf{m}^*$ essentially characterizes the set S, as all quantitative requirements of our theorem are expressed only in terms of the number of vertices, n, the number of parts, k, and $\mathbf{m}^*$.

The proof relies on a new concentration inequality we develop for symmetric subsets of the binary cube which, as we shall see, is *sharp*. Besides enabling the study of monotone properties, our results allow one to obtain tight estimates of local graph features, such as the expectation and variance of subgraph counts.

2 Motivation

As stated, our goal is to enable the study of the uniform measure over sets of graphs. The first step in this direction is to identify a "language" for specifying sets of graphs that is expressive enough to be interesting but restricted enough to be tractable.

Arguably the most natural way to introduce structure on a set is to impose symmetry. Formally this is expressed as the *invariance* of the set's characteristic function under the action of a group of transformations. In this work, we explore the progress that can be made if we define an *arbitrary* partition of the edges and take the set of transformations to be the the Cartesian product of all possible permutations of the edges (indices) within each part (symmetric group). While our work is only a first step towards a theory of extracting independence from symmetry, we argue that symmetry with respect to an edge partition is well-motivated for two reasons.

Existing Models. The first is that such symmetry, typically in a very rigid form, is already implicit in several random graph models besides $G(n, m)$. Among them are Stochastic Block Models (SBM), which assume the much stronger property of symmetry with respect to a *vertex* partition, and Stochastic Kronecker Graphs [13]. The fact that our notion of symmetry encompasses SBMs is particularly pertinent in light of the theory of Graph Limits [15], since inherent in the construction of the limiting object is an intermediate approximation of the sequence of graphs by a sequence of SBMs, via the (weak) Szemerédi Reguralitу Lemma [7, 3]. Thus, any property that is encoded in the limiting object, typically subgraph densities, is expressible within our framework.

Enabling the Expression of Geometry. A strong driving force behind the development of recent random graph models has been the incorporation of geometry, an extremely natural backdrop for network formation. Typically this is done by embedding the vertices in some (low-dimensional) metric space and assigning probabilities to edges as a function of distance. Perhaps the *most significant feature* of our work is that it fully supports the expression of geometry but in a far more light-handed manner, i.e., without imposing any specific geometric requirement. This is achieved by (i) using edge-partitions to abstract away geometry as a symmetry rendering edges of the same length equivalent, while (ii) recognizing that there exist *macroscopic* constraints on the set of feasible graphs, e.g., the total edge length. Most obviously, in a physical network where edges (wire, roads) correspond to a resource (copper, concrete) there is a bound on how much can be invested to create the network while, more generally, cost (length) may represent a number of different notions that distinguish between edges.

3 Applications

A common assumption throughout the paper is the existence of a partition of the edges, expressing prior information about the setting at hand. Two prototypical examples are: *vertex-induced partitions*, as in the SBM, and *geometry induced partitions*, as in the d-dimensional lattice (torus). The applicability of our framework depends crucially on whether the partition is fine enough to express the desired property S. The typical pipeline is: (i) translate prior information in a partition of the edges, (ii) express the set of interest S as a specification on the edge-profile $\mathbf{m}$, (iii) solve the entropy-optimization problem and obtain the matrix $\mathbf{Q}^*$, and finally, (iv) perform all analyses and computations using the product measure $G(n, \mathbf{Q}^*)$, typically exploiting results from random graph theory and concentration of measure. Below are some examples.

Budgeted Graphs. Imagine that each possible edge e has multiple attributes that can be categorical (type of relation) or operational (throughput, latecency, cost, distance), compactly encoded as vector $\mathbf{X}_e \in \mathbb{R}^d$. We can form a partition $\mathcal{P}$ by grouping together edges that have identical attributes. Let $\boldsymbol{X} = [\boldsymbol{X}_1 \dots \boldsymbol{X}_k] \in \mathbb{R}^{d \times k}$ be the matrix where we have stacked the attribute vectors from each group and $\mathbf{b}$ be a vector of budgets. In this setting we might be interested in the *affine set* of graphs $S(\boldsymbol{X}, \boldsymbol{b}) = \{G \in \mathcal{G}_n | \boldsymbol{X} \cdot \mathbf{m}(G) \leq \boldsymbol{b}\}$, which can express a wide range of constraints. For such a set, besides generality of expression, the entropy optimization problem has a closed-form analytic solution in terms of the dual variables $\boldsymbol{\lambda} \in \mathbb{R}^d_+$. The probability of an edge (u, v) in part ℓ is given by: $Q^*_{uv}(S) = [1 + \exp(\boldsymbol{X}_\ell^\intercal \boldsymbol{\lambda})]^{-1}$.

Navigability. In [11, 12], Kleinberg gave sufficient conditions for greedy routing to discover paths of poly-logarithmic length between any two vertices in a graph. One of the most general settings where such navigability is possible is set-systems, a mathematical abstraction of the relevant geometric properties of grids, regular-trees and graphs of bounded doubling dimension. The essence of navigability lies in the requirement that for any vertex in the graph, the probability of having an edge to a vertex at distance in the range $[2^{i-1}, 2^i)$, i.e., at distance scale i, is approximately uniform for all $i \in [\log n]$. In our setting, we can partition the $\binom{n}{2}$ edges according to distance scale so that part P_i includes all possible edges between vertices at distance scale i. In [1] we prove that by considering a single linear constraint where the cost of edges in scale i is proportional to i, we recover Kleinberg's results on navigability in set-systems, *without* any independence assumptions regarding network formation, or coordination between the vertices (such as using the same probability distribution). Besides establishing the robustness of navigability, eschewing a specific mechanism for (navigable) network formation allows us to recast navigability as a property of networks brought about by economical (budget) and technological (cost) advancements.

Percolation Avoidance. To show that interesting non-linear constraints can also be accommodated we focus on the case of the Stochastic Block Model. Consider a social network consisting of q groups of sizes $(\rho_1, \dots, \rho_q) \cdot n$, where $\rho_i > 0$ for $i \in [q]$. As the partition of edges is naturally induced by the partition of vertices, for simplicity we adopt a double indexing scheme and instead of the edge-profile vector $\mathbf{m} \in \mathbb{R}^{\binom{q+1}{2}}$ we are going to use a symmetric edge-progile matrix $\mathbf{M} \in \mathbb{R}^{q \times q}$. Consider the property S_ϵ that a specific group s acts as the "connector", i.e., that the graph induced by the remaining groups should have no component of size greater than ϵn for some arbitrarily small $\epsilon > 0$. While the set S_ϵ is not symmetric with respect to this partition our result can still be useful, as follows.

Using a well known connection between multitype branching processes [16] and the existence of a giant component (percolation) in mean-field models, such as $G(n,p)$ and SBM, we can cast the existence of the giant component in terms of a condition on the number of edges between each block. Concretely, given the edge-profile matrix $\mathbf{M}$, for a given cluster $s \in [q]$ define the $(q-1) \times (q-1)$ matrix: $T(\mathbf{M})_{ij} := \frac{m_{ij}}{n^2 \rho_i}, \forall i,j \in [q] \setminus \{s\}$ that encapsulates the dynamics of a multi-type branching process. Let $\|\cdot\|_2$ denote the *operator norm* (maximum singular value). A classic result of branching processes asserts that if $\|T(\mathbf{M})\|_2 < 1$ no giant component exists. Thus, in our framework, the property $S_\epsilon = \{$ no giant component without vertices from $s\}$, can be accurately approximated under the specific partition $\mathcal{P}$ by the set of graphs S for which $\mathbf{M}(S) = \{\mathbf{M} : ||T(\mathbf{M})||_2 < 1\}$, where additionally the set $\mathbf{M}(S)$ is convex as $||T(M)||_2$ is a convex function of $T(M)$ which is convex (linear) in $\mathbf{M}$.

4 Definitions and Results

We start with some notation. We will use lower case boldface letters to denote vectors and uppercase boldface letters to denote matrices. Further, we fix an arbitrary enumeration of the $N = \binom{n}{2}$ edges and sometimes represent the set of all graphs on n vertices as $H_N = \{0,1\}^N$. We will refer to an element of $x \in H_N$ interchangeably as a graph and a string. Given a partition $\mathcal{P} = (P_1, \ldots, P_k)$ of $[N]$, we define $\Pi_N(\mathcal{P})$ to be the set of all permutations acting only within blocks of the partition.

Edge Block Symmetry. *Fix a partition $\mathcal{P}$ of $[N]$. A set $S \subseteq H_N$ is called $\mathcal{P}$-symmetric if it is invariant under the action of $\Pi_N(\mathcal{P})$. Equivalently, if $\mathbb{I}_S(x)$ is the indicator function of set S, then $\mathbb{I}_S(x) = \mathbb{I}_S(\pi(x))$ for all $x \in H_N$ and $\pi \in \Pi_N(\mathcal{P})$.*

The number of parts $k = |\mathcal{P}|$ gives a rough indication of the amount of symmetry present. For example, when $k = 1$ we have maximum symmetry as all edges are equivalent. In a stochastic block model (SBM) with ℓ vertex classes we have $k = \binom{\ell}{2}$. For a d-dimensional lattice, partitioning the $\binom{n}{2}$ edges by distance results in roughly $k = n^{1/d}$ parts. Finally, if $k = N$ there is no symmetry whatsoever. Our results accommodate partitions with as many as $O(n^{1-\epsilon})$ parts. This is way more than enough for most situations. For example, as we just saw, in d-dimensional lattices there are $O(n^{1/d})$ distances. Generically, if we have n points such that the nearest pair habe distance 1 and the farthest have distance D, fixing any $\delta > 0$ and binning together all edges of length $[(1+\delta)^i, (1+\delta)^{i+1})$ for $i \geq 0$, yields only $O(\delta^{-1} \log D)$ classes.

Recall that given a partition $\mathcal{P} = (P_1, \ldots, P_k)$ of H_N and a graph $x \in H_N$, the edge profile of x is $\mathbf{m}(x) := (m_1(x), \ldots, m_k(x))$, where $m_i(x)$ is the number of edges of x from P_i, and that the image of a $\mathcal{P}$-symmetric set S under $\mathbf{m}$ is denoted as $\mathbf{m}(S) \subseteq \mathbb{R}^k$. The edge-profile is crucial to the study of $\mathcal{P}$-symmetric sets due to the following intuitively obvious fact.

Proposition 1. *Any function $f : H_N \to \mathbb{R}$ invariant under $\Pi_N(\mathcal{P})$ depends only on the edge-profile $\mathbf{m}(\mathbf{x})$.*

Definition 1. *Let* $p_i = |P_i|$ *denote the number of edges in part* i *of partition* $\mathcal{P}$.

Edge Profile Entropy. *Given an edge profile* $\mathbf{v} \in \mathbf{m}(S)$ *define the entropy of* $\mathbf{v}$ *as* $\textsc{Ent}(\mathbf{v}) = \sum_{i=1}^{k} \log \binom{p_i}{v_i}$.

Using the edge-profile entropy we can express the induced distribution on $\mathbf{m}(S)$ as $\mathbb{P}(\mathbf{v}) = \frac{1}{|S|} e^{\textsc{Ent}(\mathbf{v})}$. The crux of our argument is now this: the only genuine obstacle to S being approximable by a product measure is degeneracy, i.e., the existence of multiple, well-separated edge-profiles that maximize $\textsc{Ent}(\mathbf{v})$. The reason we refer to this as degeneracy is that it typically encodes a hidden symmetry of S with respect to $\mathcal{P}$. For example, imagine that $\mathcal{P} = (P_1, P_2)$, where $|P_1| = |P_2| = p$, and that S contains all graphs with $p/2$ edges from P_1 and $p/3$ edges from P_2, or vice versa. Then, the presence of a single edge $e \in P_i$ in a uniformly random $G \in S$ boosts the probability of all other edges in P_i, rendering a product measure approximation impossible.

Note that since $\mathbf{m}(S)$ is a discrete set, it is non-trivial to quantify what it means for the maximizer of $\textsc{Ent}$ to be "sufficiently unique". For example, what happens if there is a unique maximizer of $\textsc{Ent}(\mathbf{v})$ strictly speaking, but sufficiently many near-maximizers to potentially receive, in aggregate, a majority of the measure? To strike a balance between conceptual clarity and generality we focus on the following.

Convexity. *Let* $\mathrm{Conv}(A)$ *denote the convex hull of a set* A. *Say that a* $\mathcal{P}$-*symmetric set* $S \subseteq \mathcal{G}_N$ *is convex iff the convex hull of* $\mathbf{m}(S)$ *contains no new integer points, i.e., if* $\mathrm{Conv}(\mathbf{m}(S)) \cap \mathbb{N}^k = \mathbf{m}(S)$.

Let $H_{\mathcal{P}}(\mathbf{v})$ be the approximation to $\textsc{Ent}(\mathbf{v})$ that results by replacing each binomial term with its binary entropy approximation via the first term in Stirling's approximation.

Entropic Optimizer. *Let* $\mathbf{m}^* = \mathbf{m}^*(S) \in \mathbb{R}^k$ *be the solution to* $\max_{\mathbf{v} \in \mathrm{Conv}(\mathbf{m}(S))} H_{\mathcal{P}}(\mathbf{v})$.

Defining the optimization over the convex hull of $\mathbf{m}(S)$ will allow us to study the set S by studying only the properties of the maximizer $\mathbf{m}^*$. Clearly, if a $\mathcal{P}$-symmetric set S has entropic optimizer $\mathbf{m}^* = (m_1^*, \ldots, m_k^*)$, the natural candidate product measure for each $i \in [k]$ assigns probability m_i^*/p_i to all edges in part P_i. The challenge is to relate this product measure to the uniform measure on S by proving concentration of the induced measure on $\mathbf{m}(S)$ around a point near $\mathbf{m}^*$. For that we need (i) the vector $\mathbf{m}^*$ to be "close" to a vector in $\mathbf{m}(S)$, and (ii) to control the decrease in entropy "away" from $\mathbf{m}^*$. To quantify this second notion we need the following parameters, expressing the geometry of convex sets.

Definition 2. *For a $\mathcal{P}$-symmetric convex set S define*

$$\text{Thickness:} \qquad \mu = \mu(S) = \min_{i \in [k]} \min\{m_i^*, p_i - m_i^*\} \tag{2}$$

$$\text{Condition number:} \qquad \lambda = \lambda(S) = \frac{5k \log n}{\mu(S)} \tag{3}$$

$$\text{Resolution:} \qquad r = r(S) = \frac{\lambda + \sqrt{\lambda^2 + 4\lambda}}{2} > \lambda \tag{4}$$

The most important of the above three parameters is *thickness*. Its role is to quantify how close the optimizer $\mathbf{m}^*(S)$ comes, in any coordinate, to the natural boundary $\{0, p_1\} \times \ldots \times \{0, p_k\}$, where the entropy of a class becomes zero. As a result, thickness determines the *coordinate-wise concentration* around the optimum.

The *condition number* $\lambda(S)$, on the other hand, quantifies the robustness of S. To provide intuition, in order for the product measure approximation to be accurate for every class of edges (part of $\mathcal{P}$), fluctuations in the number of edges of order $\sqrt{m_i^*}$ need to be "absorbed" in the mean m_i^*. For this to happen with polynomially high probability for a single part, standard results imply we must have $m_i^* = \Omega(\log(n))$. We absorb the dependencies between parts by taking a union bound, thus multiplying by the number of parts, yielding the numerator in (3). Our results give strong probability bounds when $\lambda(S) \ll 1$, i.e., when in a typical graph in S the number of edges from each part P_i is $\Omega(k \log n)$ edges away from triviality, i.e., both from 0 and from $|P_i| = p_i$, a condition we expect to hold in all natural applications. We can now state our main result.

Theorem 2 (Main result). *Let $\mathcal{P}$ be any edge-partition and let S be any $\mathcal{P}$-symmetric convex set. For every $\epsilon > \sqrt{12\lambda(S)}$, the uniform measure on S is (ϵ, δ)-sandwichable, where $\delta = 2 \exp\left[-\mu(S)\left(\frac{\epsilon^2}{12} - \lambda(S)\right)\right]$.*

Remark 1. As a sanity check we see that as soon as $m \gg \log n$, Theorem 2 recovers the sandwichability of $G(n, m)$ by $G(n, p(m))$ as sharply as the Chernoff bound, up to the constant factor $\frac{1}{12}$ in the exponent.

Theorem 2 follows by analyzing the natural coupling between the uniform measure on S and the product measure corresponding to the entropic optimizer $\mathbf{m}^*$. Our main technical contribution is Theorem 3 below, a concentration inequality for $\mathbf{m}(S)$ when S is a convex symmetric set. The *resolution*, $r(S)$, defined in (4) above, reflects the narrowest *concentration interval* that can be proved by our theorem. When $\lambda(S) \ll 1$, as required for the theorem to be meaningfully applied, it scales optimally as $\sqrt{\lambda(S)}$.

Theorem 3. *Let $\mathcal{P}$ be any edge-partition, let S be any $\mathcal{P}$-symmetric convex set, and let $\mathbf{m}^*$ be the entropic optimizer of S. For all $\epsilon > r(S)$, if $G \sim U(S)$, then*

$$\mathbb{P}_S\left(|\mathbf{m}(G) - \mathbf{m}^*| \le \epsilon \tilde{\mathbf{m}}^*\right) \ge 1 - \exp\left(-\mu(S)\left(\frac{\epsilon^2}{1+\epsilon} - \lambda(S)\right)\right) \quad , \tag{5}$$

where $\mathbf{x} \le \mathbf{y}$ means that $x_i \le y_i$ for all $i \in [k]$, and $\tilde{m}_i = \min\{m_i^*, p_i - m_i^*\}$.

The intuition driving concentration is that as *thickness* increases two phenomena occur: (i) vectors close to $\mathbf{m}^*$ capture a larger fraction of the measure, and (ii) the decay in entropy away from $\mathbf{m}^*$ becomes steeper. These joint forces compete against the probability mass captured by vectors "away" from the optimum. The point were they prevail corresponds to $\lambda(S) \ll 1$ or, equivalently, $\mu(S) \gg 5k\log(n)$. Assuming $\lambda(S) \ll 1$ the probability bounds we give scale as $n^{-\Omega(k\epsilon^2)}$. Without assumptions on S, and up to the constant 5 in (3), this is *sharp*, per Proposition 2 below.

5 Technical Overview

In this section, we present an overview of the technical work involved in proving Theorems 2 and 3. Most of the work lies in the concentration result, Theorem 3.

Concentration. The general idea is to identify a high-probability subset $\mathcal{L} \subseteq \mathbf{m}(S)$ by integrating the probability measure around the entropy-maximizing profile $\mathbf{m}^*$. Since ultimately our goal is to couple the uniform measure with a product measure, we need to establish concentration for the number of edges from each and every part, i.e., in every coordinate. There are two main issues: (i) we do not know $|S|$, and (ii) we must quantify the decrease in entropy as a function of the L_∞ distance from the maximizer $\mathbf{m}^*$. Our strategy to address these issues is:

Size of S. We bound $\log|S|$ from below by the contribution to $\log|S|$ of the entropic optimal edge-profile $\mathbf{m}^*$, thus upper-bounding the probability of every $\mathbf{v} \in \mathbf{m}(S)$ as

$$\log \mathbb{P}_S(\mathbf{v}) = \textsc{Ent}(\mathbf{v}) - \log(|S|) \leq \textsc{Ent}(\mathbf{v}) - \textsc{Ent}(\mathbf{m}^*) \ . \tag{6}$$

This is the crucial step that opens up the opportunity of relating the probability of a vector $\mathbf{v}$ to the distance $\|\mathbf{v} - \mathbf{m}^*\|_2$ through analytic properties of entropy. Key to this is the definition of $\mathbf{m}^*$ as the maximizer over $\mathrm{Conv}(\mathbf{m}(S))$ instead of over $\mathbf{m}(S)$.

Proposition 2. *If $S = \mathcal{G}_n$ and $\mathcal{P}$ is any k-partition such that $|P_i| = \binom{n}{2}/k$ for all i, then $\log(|S|) - \textsc{Ent}(\mathbf{m}^*) = \Omega(k\log(n))$.*

Proposition 2 demonstrates that unless one utilizes specific geometric properties of the set S enabling integration around $\mathbf{m}^*$, instead of using a point-bound for $\log|S|$, a loss of $\Omega(k\log(n))$ is unavoidable. In other words, either one makes more assumptions on S besides symmetry and "convexity", or the claimed error term is optimal.

Distance bounds: To bound from below the rate at which entropy decays as a function of the component-wise distance from the maximizer $\mathbf{m}^*$, we first approximate $\textsc{Ent}(\mathbf{v})$ by $H_{\mathcal{P}}(\mathbf{v})$ (the binary entropy introduced earlier) to get a smooth function. Then, exploiting the separability, concavity and differentiability of binary entropy, we obtain component-wise distance bounds using a second-order Taylor approximation. At this step we also lose a cumulative factor of order $3k \log n$ stemming from Stirling approximations and the subtle point that the maximizer $\mathbf{m}^*$ might not be an integer point. The constant 3 can be improved, but in light of Proposition 2 this would be pointless and complicate the proof unnecessarily.

Union bound: Finally, we integrate the obtained bounds outside the set of interest by showing that even if all "bad" vectors where placed right at the boundary of the set, where the lower bound on the decay of entropy is smallest, the total probability mass would be exponentially small. The loss incurred at this step is of order $2k \log n$, since there are at most n^{2k} bad vectors.

Relaxing Conclusions. Our theorem seeks to provide concentration simultaneously for all parts. That motivates the definition of thickness parameter $\mu(S)$ as the minimum distance from the trivial boundary that any part has at the optimum $\mathbf{m}^*$. Quantifying everything in terms of $\mu(S)$ is a very conservative requirement. For instance, if we define the set S to have no edges in a particular part of the partition, then $\mu(S)$ is 0 and our conclusions become vacuous. Our proofs in reality generalize, to the case where we confine our attention only to a subset $I \subseteq [k]$ of blocks in the partition. In particular, if one defines I^* as the set of parts whose individual *thickness* parameter $\tilde{m}_i = \min\{m_i, p_i - m_i\}$ is greater than $5k \log n$, both theorems hold for the subset of edges $\cup_{i \in I^*} P_i$. In essence that means that for every part that is "well-conditioned", we can provide concentration of the number of edges and approximate monotone properties of only those parts by coupling them with product measures.

Relaxing Convexity. Besides partition symmetry, that comprises our main premise and starting point, the second main assumption made about the structure of S is *convexity.* In the proof convexity is used only to argue that: (i) the maximizer $\mathbf{m}^*$ will be close to some vector in $\mathbf{m}(S)$, and (ii) that the first order term in the Taylor approximation of the entropy around $\mathbf{m}^*$ is always negative. Since the optimization problem is defined on the convex hull of $\mathbf{m}(S)$, the convexity of $\mathrm{Conv}(\mathbf{m}(S))$ implies (ii), independently of whether $\mathbf{m}(S)$ is convex or not. We thus see that we can replace convexity of $\mathcal{P}$-symmetric sets with *approximate unimodality.*

Definition 3. *A $\mathcal{P}$-symmetric set S is called Δ-unimodal if the solution $\mathbf{m}^*$ to the entropy optimization problem defined in Section 2, satisfies:*

$$d_1(\mathbf{m}^*, S) := \min_{\mathbf{v} \in \mathbf{m}(S)} \|\mathbf{m}^* - \mathbf{v}\|_1 \leq \Delta \tag{7}$$

Convexity essentially implies that the set S is k-unimodal as we need to round each of the k coordinates of the solution to the optimization problem to the nearest integer. Under this assumption, all our results apply by only changing the condition number of the set to $\lambda(S) = \frac{(2\Delta + 3k) \log n}{\mu(S)}$. In this extended abstract, we opted to present our results by using the familiar notion of convexity to convey intuition on our results and postpone the presentation in full generality for the full version of the paper.

Coupling. To prove Theorem 2 using our concentration result, we argue as follows. Conditional on the edge-profile, we can couple the generation of edges in different parts independently, in each part the coupling being identical to that between $G(n,m)$ and $G(n,p)$. Then, using a union bound we can bound the probability that all couplings succeed, given an appropriate $\mathbf{v}$. Finally, using the concentration theorem we show that sampling an appropriate edge-profile $\mathbf{v}$ happens with high probability.

References

1. Dimitris Achlioptas and Paris Siminelakis. Navigability is a robust property. *CoRR*, abs/1501.04931, 2015.
2. Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.
3. C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao. An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions. *ArXiv e-prints*, January 2014.
4. Fan RK Chung and Linyuan Lu. *Complex graphs and networks*, volume 107. American mathematical society Providence, 2006.
5. Richard Durrett. *Random graph dynamics*, volume 20. Cambridge university press, 2007.
6. Paul Erdős and Alfréd Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.
7. Alan Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.
8. Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, pages 1141–1144, 1959.
9. Anna Goldenberg, Alice X. Zheng, Stephen E. Fienberg, and Edoardo M. Airoldi. A survey of statistical network models. *Found. Trends Mach. Learn.*, 2(2):129–233, February 2010.
10. Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.
11. Jon M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845, August 2000.
12. Jon M. Kleinberg. Small-world phenomena and the dynamics of information. In Thomas G. Dietterich, Suzanna Becker, and Zoubin Ghahramani, editors, *NIPS*, pages 431–438. MIT Press, 2001.
13. Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. Kronecker graphs: An approach to modeling networks. *The Journal of Machine Learning Research*, 11:985–1042, 2010.
14. Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.
15. László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.
16. C.J. Mode. *Multitype branching processes: theory and applications*. Modern analytic and computational methods in science and mathematics. American Elsevier Pub. Co., 1971.
17. Mark Newman. *Networks: an introduction*. Oxford University Press, 2010.