Symmetric Graph Properties Have Independent Edges

Dimitris Achlioptas[a,b,1,2], Paris Siminelakis[c,1,3,*]

[a] *Department of Computer Science University of California, Santa Cruz*
[b] *Department of Informatics & Telecommunications, University of Athens*
[c] *Department of Electrical Engineering, Stanford University*

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?*

Keywords: Random Graphs, Symmetric Graph Properties Concentration of Measure, Approximate Independence

*Corresponding author

Email addresses: optas@cs.ucsc.edu (Dimitris Achlioptas), psimin@stanford.edu (Paris Siminelakis)

[1]Research supported by ERC Starting Grant StG-210743.

[2]Research supported by NSF Grant CCF-1514128, an Alfred P. Sloan Fellowship, 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.

[3]Supported by an Onassis Foundation Scholarship.

1. Introduction

Since their introduction in 1959 by Erdős and Rényi [1] and Gilbert [2], respectively, $G(n, m)$ and $G(n, p)$ random graphs have dominated the mathematical study of random networks [3, 4]. 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 [3]. 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.

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 [5]. 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 [6] 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 [7] and for a more mathematical treatment to the books of Chung and Lu [8] and of Durrett [9].

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 $m \sim \text{Bin}\left(\binom{n}{2}, p\right)$ 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 $\mathbb{P}[G^- \subseteq G \subseteq G^+] \geq 1-\delta$.*

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

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 4), 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 while remaining amenable to approximability by a product measure.

We state our results formally in Section 3. 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 (Section 7). This is achieved by proving that for certain sets S the probabilities of local events are close to the probabilities assigned by the product measure within a vanishing multiplicative constant.

Outline. The paper is organized as follows. In the next section we provide motivation about partition symmetry and its connections to contemporary work on random graphs. In Section 3, we give formal definitions and state our results. Section 4 is devoted to providing a high-level technical summary of the proofs, which are provided in Sections 5 and 6. Section 7, provides our result about probabilities of local events. Finally, in Section 8 we discuss applications of our theorems.

2. Motivation

As stated, our goal is to make the study of the uniform measure over sets of graphs easier. The first step towards this goal 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 symmetry with respect to a *vertex* partition, i.e., the very special case in which the edge partition factorizes over the vertices, and Stochastic Kronecker Graphs [10]. The fact that our notion of symmetry encompasses SBMs is particularly pertinent in light of the theory of Graph Limits [11], 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 Regularity Lemma [12, 13]. 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. This is achieved by (i) using edge-partitions to abstract away geometry as a symmetry rendering certain edges equivalent, while (ii) recognizing that there exist *macroscopic* constraints on the set of feasible graphs, e.g., total edge length. Most obviously, in physical networks where edges consume a resource (copper, concrete) there is a bound on how much can be invested to create the network while,

more generally, cost may represent several different notions that distinguish between edges.

3. Definitions and Results

Notation. We use lower case boldface letters to denote vectors and uppercase boldface letters to denote matrices. We adopt the usual asymptotic notation e.g. $f(n) = O(g(n))$ to denote that there is a $n_0 \in \mathbb{N}$ and a constant $C > 0$ such that $f(n) \leq Cg(n)$ for all $n \geq n_0$. Throughout the paper, we denote with $\mathbb{P}_S(A)$ the probability of the event A under the uniform measure on the set S and more generally for a probability measure ν, we write $\mathbb{P}_\nu(A)$ to denote the same probability under ν. Further, we say that an event A holds with high probability (w.h.p.) when $\mathbb{P}(A) \geq 1 - o(1)$. Additionally, we fix an arbitrary enumeration of the $N = \binom{n}{2}$ edges and interchangeably represent the set of all graphs on n vertices as $H_N = \{0,1\}^N$ and $\mathcal{G}_n$. In the same spirit will refer to an element of $\mathbf{x} \in H_N$ interchangeably as a graph and a string.

Partition Symmetry. *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. 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 have 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 $\mathbf{m}(\mathbf{x})$.*

Proof. Fix an $\mathbf{x} \in H_n$ and consider the set $O(\mathbf{x}) := \{\mathbf{y} \in H_n : \exists \pi \in \Pi_n(\mathcal{P})$ such that $y = \pi(\mathbf{x})\}$ and call it the *orbit* of $\mathbf{x}$ under $\Pi_n(\mathcal{P})$(note that by group property orbits form a partition of H_n). The assumption of symmetry, implies that f is constant for all $\mathbf{y} \in O(\mathbf{x})$:

$$f(\mathbf{y}_1) = f(\mathbf{y}_2) = f(\mathbf{x}),\ \forall \mathbf{y}_1, \mathbf{y}_2 \in O(\mathbf{x})$$

By definition of $\Pi_n(\mathcal{P})$, for any $\mathbf{x} \in H_n$ there is a permutation $\pi_x \in \Pi_n(\mathcal{P})$, such that i) $\pi_x(\mathbf{x}) = (\pi_{x,1}(x_{P_1}), \ldots, \pi_{x,k}(x_{P_k})) \in O(\mathbf{x})$, ii) for all $i \in [k]$, $\pi_{x,i}(x_{P_i})$ is a bit-string where all 1's appear consequently starting from the first position. Let as identify with each orbit $O \subset H_n$ such a distinct element $\mathbf{x}_o$. As the function of f is constant along each orbit, its value depends only through $\mathbf{x}_o$, which in turn depends only on the number of 1's(edges) in each part, encoded in the edge profile $\mathbf{m} = (m_1, \ldots, m_k)$. □

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 $H_{\mathcal{P}}(\mathbf{v}) := -\sum_{i=1}^{k} \left[v_i \log\left(\frac{v_i}{p_i}\right) + (p_i - v_i) \log\left(\frac{p_i - v_i}{p_i}\right)\right]$ be the approximation to $\textsc{Ent}(\mathbf{v})$ that results by replacing each binomial term with its first order 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 salient properties of the set of interest.

Definition 2. *For a* $\mathcal{P}$*-symmetric 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 the minimum coordinate-wise distance of the optimizer $\mathbf{m}^*(S)$ 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, S be any $\mathcal{P}$-symmetric convex set, let $\mathbf{m}^*$ be the entropic optimizer of S. For all $\epsilon > r(S)$, if $G \sim U(S)$, then*

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

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.

4. 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 "ball" $\mathcal{B} \subseteq \mathbf{m}(S)$ around the entropy-maximizing profile $\mathbf{m}^*$ such that the remaining set $\mathcal{B}^c$ has negligible probability. 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

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

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 binary entropy $H_{\mathcal{P}}$. Key to this is the definition of $\mathbf{m}^*$ as the maximizer over $\text{Conv}(\mathbf{m}(S))$ instead of over $\mathbf{m}(S)$. An error term appears at this point due to Stirling approximation of binomial coefficients as well as due to the fact that $\mathbf{m}^*$ might not be in $\mathbf{m}(S)$ and thus we need to add an error term caused by "rounding" $\mathbf{m}^*$ to a point in $\mathbf{m}(S)$. We deal with the latter issue in Proposition 3 and show that the error term can be bounded by a term of the order $O(k\log(n))$. The following proposition demonstrates that unless one utilizes additional 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.

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* $\log(|S|) - \textsc{Ent}(\mathbf{m}^*) = \Omega(k \log(n))$.

Proof. Consider $\mathcal{P}$ any balanced partition consisting of k-parts and let $S = \mathcal{G}_n$ to be the space of all graphs. Then $|S| = 2^{\binom{n}{2}}$ and $\mathbf{m}^*(S)$ is the all $\binom{n}{2}/2k$ vector (all blocks half full). Using Stirling's approximation of the factorial we have that:

$$\begin{aligned} \log(|S|) - \textsc{Ent}(\mathbf{m}^*) &\geq \binom{n}{2} \log 2 - 2k \log \binom{\binom{n}{2}/k}{\frac{1}{2}\binom{n}{2}/k} && (7) \\ &\geq k \log n - \frac{k}{2} \log k && (8) \end{aligned}$$

For $k < n^2$ the last expression is always $\Omega(k \log(n))$. □

Distance bounds. To bound the rate at which entropy decays as a function of the component-wise distance from the maximizer $\mathbf{m}^*$ from below, 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 (Lemma 2) using a second-order Taylor approximation . At this step we also lose a cumulative factor of order $3k \log n$ stemming from Stirling approximations (Lemma 1) and the subtle point that the maximizer $\mathbf{m}^*$ might not be an integer point (Proposition 3). 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:*

$$d_1(\mathbf{m}^*, S) := \min_{\mathbf{v} \in \mathbf{m}(S)} \|\mathbf{m}^* - \mathbf{v}\|_1 \leq \Delta \tag{9}$$

Convexity of $\mathbf{m}(S)$ implies that the set S is k-unimodal (Proposition 3) as we only need to round each of the k coordinates of the solution to the optimization problem to an integer. Under this assumption, all our results apply by only changing the *condition number* of the set to $\lambda(S) = \frac{(2\Delta+3k)\log n}{\mu(S)}$.

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.

5. Proof of Theorem 3

In this section we prove Theorem 3. For the purposes of the proof we are going to employ, instead of $\mathbf{m}(x)$, a different parametrization in terms of the

probability-profile $\mathbf{a}(x) = (a_1(x), \ldots, a_k(x)) \in [0,1]^k$ where $a_i(x) = m_i(x)/p_i$. This will be convenient both in calculations as well as conceptually as $a_i(x)$ represents the effective edge density of a part P_i in the partition. We start by approximating the entropy of an edge-profile via the $\mathcal{P}$-entropy.

Definition 4. *Given a partition $\mathcal{P}$, the $\mathcal{P}$-entropy of $\mathbf{a} \in [0,1]^k$ is*

$$H_{\mathcal{P}}(\mathbf{a}) := -\sum_{i=1}^{k} p_i \left[a_i \log a_i + (1-a_i)\log(1-a_i)\right] \tag{10}$$

The $\mathcal{P}$-entropy is simply the entropy of the product measure defined over edges through $\mathbf{a}$. We slightly abuse the notation and also define the $\mathcal{P}$-entropy as before in terms of the edge-profile:

$$H_{\mathcal{P}}(\mathbf{v}) := -\sum_{i=1}^{k} \left[v_i \log\left(\frac{v_i}{p_i}\right) + (p_i - v_i)\log\left(\frac{p_i - v_i}{p_i}\right)\right] \tag{11}$$

Let $\mathcal{M}_{\mathcal{P}} := \{0, \ldots, p_1\} \times \ldots \times \{0, \ldots, p_k\}$ be the space of all possible vectors $\mathbf{m}$. In what follows we sometimes suppress the dependence of the quantities in $\mathbf{m}$ or $\mathbf{a}$ to ease the notation.

Lemma 1. *Let $\mathbf{m} \in \mathcal{M}_{\mathcal{P}}$ be an edge-profile and $\mathbf{a} \in [0,1]^k$ be the corresponding probability profile, then:*

$$\textsc{Ent}(\mathbf{m}) = \mathcal{H}_{\mathcal{P}}(\mathbf{a}) - \gamma(n)$$

where $0 \le \gamma(n) \le k \log n$ tends to 0 as m_i and $p_i - m_i$ tend to infinity.

Proof. We begin by providing the first order Stirling approximation for a single term of the form $\log \binom{p_i}{m_i}$. Specifically, since $m_i = p_i a_i$ and by using $\log n! = n \log n - n + \frac{1}{2}\log n + \theta_n$, where $\theta_n \in (0,1]$, we get:

$$\begin{aligned} \log \binom{p_i}{m_i} &= \log(p_i!) - \log(m_i!) - \log((p_i - m_i)!) \\ &= -p_i \left[a_i \log a_i + (1-a_i)\log(1-a_i)\right] - \delta_n(a_i, p_i) \ , \end{aligned}$$

where $0 \le \delta_n(a_i, p_i) \le \log n$. Summing the derived expression for all $i \in [k]$ gives:

$$\begin{aligned} \textsc{Ent}(\mathbf{m}) &= -\sum_{i=1}^{k} p_i \left[a_i \log a_i + (1-a_i)\log(1-a_i)\right] - \sum_{i=1}^{k} \delta_n(a_i, p_i) \\ &= H_{\mathcal{P}}(\mathbf{a}) - \gamma(n) \ , \end{aligned}$$

where $0 \le \gamma(n) \le k \log n$. □

Proposition 3. *Let $I := \{i \in [k] : m_i^* \notin \{0, p_i\}\}$ denote the set of coordinates of $\mathbf{m}^*$ that are not on the trivial boundary. There exists an edfe profile $\mathbf{z}^* \in \mathbf{m}(S)$ such that $H_{\mathcal{P}}(\mathbf{m}^*) - H_{\mathcal{P}}(\mathbf{z}^*) \leq |I|(2\log(n) - \log(1-\theta))$.*

Proof. Since, $\mathbf{m}^*$ is in the convex hull of $\mathbf{m}(S)$, convexity ensures that the vector $\mathbf{z}^*$, defined by rounding coordinates of $\mathbf{m}^*$ close to 0 to 1 and coordinates close to p_i to $p_i - 1$, is itself in $\mathbf{m}(S)$. Further, letting $\mathbf{m}_\theta := \theta\mathbf{m}^* + (1-\theta)\mathbf{z}$, the Mean Value Theorem implies that there is $\theta \in (0,1)$ such that

$$H_{\mathcal{P}}(\mathbf{m}^*) - H_{\mathcal{P}}(\mathbf{z}) = \nabla H_{\mathcal{P}}(\mathbf{m}_\theta)^\intercal(\mathbf{m}^* - \mathbf{z}) \leq |I| \max_{i \in I} \left| \log\left(\frac{(m_\theta)_i}{p_i - (m_\theta)_i}\right)\right| \ .$$

In any case, $\max_{i\in I}\left|\log\left(\frac{(m_\theta)_i}{p_i-(m_\theta)_i}\right)\right| \leq \max_{i\in I}\log\left(\frac{p_i-1+\theta}{1-\theta}\right) \leq \max_{i\in I}\log(p_i) - \log(1-\theta) \leq 2\log(n) - \log(1-\theta)$, where we used the trivial bound $p_i \leq n^2$. □

Next, using the Taylor remainder theorem and the connection with $\mathcal{P}$-entropy, we obtain geometric estimates on the decay of entropy around $\mathbf{m}^*$.

Theorem 4 (Taylor Remainder Theorem)**.** *Assume that f and all its partial derivatives are differentiable at every point of an open set $S \subseteq \mathbb{R}^k$. If $\mathbf{a}, \mathbf{b} \in S$ are such that the line segment $L(\mathbf{a}, \mathbf{b}) \subseteq S$, then there exists a point $\mathbf{z} \in L(\mathbf{a}, \mathbf{b})$ such that:*

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{a})^T(\mathbf{b} - \mathbf{a}) + \frac{1}{2}(\mathbf{b} - \mathbf{a})^T \nabla^2 f(\mathbf{z})(\mathbf{b} - \mathbf{a}) \ . \tag{12}$$

Lemma 2 (L_2 distance bounds)**.** *If $\mathbf{m}^*$ is the unique maximizer and $\mathbf{v} \in \mathrm{Conv}(\mathbf{m}(S))$, then*

$$\textsc{Ent}(\mathbf{v}) - \log|S| \leq -\sum_{i=1}^{k} \frac{(v_i - m_i^*)^2}{\max\{\tilde{m}_i^*, \tilde{v}_i\}} + 3k\log n \ , \tag{13}$$

where $\mathbf{m}_i^* = \min\{m_i^*, p_i - m_i^*\}$ (respectively $\tilde{v}_i$), is the thickness *of part $i \in [k]$.*

Proof. We start by lower bounding $\log|S|$ by the entropy of the vector $\mathbf{z}^*$ given by Proposition 3. Next, invoking Lemma 1, we rewrite the difference in entropy as a difference in $\mathcal{P}$-entropy, and finally use Proposition 3 to replace $\mathbf{z}^*$ with $\mathbf{m}^*$:

$$\begin{aligned}
\textsc{Ent}(\mathbf{v}) - \log|S| &\leq \textsc{Ent}(\mathbf{v}) - \textsc{Ent}(\mathbf{z}^*) && (14)\\
&\leq H_{\mathcal{P}}(\mathbf{v}) - H_{\mathcal{P}}(\mathbf{z}^*) + \gamma(n) && (15)\\
&\leq H_{\mathcal{P}}(\mathbf{v}) - H_{\mathcal{P}}(\mathbf{m}^*) + 3k\log n && (16)
\end{aligned}$$

The "error" term $3k \log n$ can be reduced using higher order Stirling approximations but we avoid doing so since an error of order $k \log(n)$ is unavoidable due to the approximation of $\log |S|$ by the entropy of the maximizer. Convexity of the domain $\mathrm{Conv}(\mathbf{m}(S))$ and differentiability of the $\mathcal{P}$-entropy provide the necessary conditions to use the Taylor Remainder Theorem. To make this derivation more natural, we are going to use the reparametrization in terms of the probability profiles $\mathbf{a}^*$ of $\mathbf{m}^*$ and $\mathbf{b}$ of $\mathbf{v}$.

Let $\mathbf{h}$ be a point in the linear segment $L(\mathbf{a}^*, \mathbf{b})$. We proceed with writing the expressions for partial derivatives of $H_{\mathcal{P}}$.

$$\begin{aligned} \partial_i H_{\mathcal{P}}(\mathbf{a}^*) &= -p_i \log\left(\frac{a_i^*}{1-a_i^*}\right) && (17) \\ \partial^2_{ii} H_{\mathcal{P}}(\mathbf{h}) &= -p_i \left(\frac{1}{1-h_i} + \frac{1}{h_i}\right) \ , && (18) \end{aligned}$$

while $\partial^2_{ij} f = 0$ for $i \neq j$ due to separability of the function $H_{\mathcal{P}}$. The Taylor Remainder formula, now reads:

$$H_{\mathcal{P}}(\mathbf{b}) - H_{\mathcal{P}}(\mathbf{a}^*) = \nabla H_{\mathcal{P}}(\mathbf{a}^*) \cdot (\mathbf{b} - \mathbf{a}^*) - \sum_{i=1}^{k} p_i (b_i - a_i^*)^2 \left(\frac{1}{1-h_i} + \frac{1}{h_i}\right) \quad (19)$$

Since, $\mathbf{a}^*$ is the unique solution to the MaxEnt problem and the domain is convex, the first term in the above formula is always bounded above by zero. Otherwise, there would be a direction $\mathbf{u}$ and a small enough parameter $\epsilon > 0$ such that $\mathbf{a}^* + \epsilon \mathbf{u}$ has greater entropy, a contradiction. To bound the second sum from above, let $\tilde{h}_i = \min\{h_i, 1 - h_i\}$(expressing the fact that binary entropy is symmetric around $1/2$) and use the trivial bound $\tilde{h}_i \leq \max\{\tilde{a}_i^*, \tilde{b}_i\}$. Thus,

$$H_{\mathcal{P}}(\mathbf{b}) - H_{\mathcal{P}}(\mathbf{a}^*) \leq -\sum_{i=1}^{k} p_i (b_i - a_i^*)^2 \frac{1}{\tilde{h}_i} \leq -\sum_{i=1}^{k} p_i \frac{(b_i - a_i^*)^2}{\max\{\tilde{a}_i^*, \tilde{b}_i\}} \ . \quad (20)$$

Dividing and multiplying by p_i, and writing $\tilde{v}_i = p_i \tilde{b}_i$, $\tilde{m}_i^* = p_i \tilde{a}_i^*$, gives:

$$H_{\mathcal{P}}(\mathbf{b}) - H_{\mathcal{P}}(\mathbf{a}^*) \leq -\sum_{i=1}^{k} \frac{(v_i - m_i^*)^2}{\max\{\tilde{m}_i^*, \tilde{v}_i\}} \ . \quad (21)$$

where $\mathbf{v}$ and $\mathbf{m}^*$ are the edge profiles. □

In preparation of performing the "union bound", we prove that:

Proposition 4. *The number of distinct edge-profiles* $|\mathbf{m}(S)|$ *is bounded by* $|\mathcal{M}_{\mathcal{P}}| \leq e^{2k \log n}$.

Proof. By definition, $|\mathbf{m}(S)| \leq |\mathcal{M}_{\mathcal{P}}| = \prod_{i=1}^{k}(p_i + 1) \leq (n^2)^k$. □

We are now in a position to complete the proof of the theorem.

Proof of Theorem 3. Our goal is to use the developed machinery to control the probability of deviations from the optimum at scale $\epsilon > r(S)$. Define the set $\mathcal{B}_\epsilon(\mathbf{m}^*) := \{\mathbf{v} \in \mathbf{S} : |\mathbf{v} - \mathbf{m}^*| \leq \epsilon \tilde{\mathbf{m}}^*\}$. We are going to show that $\mathbb{P}_S(\mathcal{B}^c_\epsilon(\mathbf{m}^*)) \to 0$ "exponentially" fast and thus provide localization of the edge profile within a scale ϵ for each coordinate. To that end, we write:

$$\mathbb{P}_S(\mathcal{B}^c_\epsilon(\mathbf{m}^*)) = \sum_{\mathbf{v} \in \mathcal{B}^c_\epsilon(\mathbf{m}^*)} \mathbb{P}_S(\mathbf{v}) = \sum_{\mathbf{v} \in \mathcal{B}^c_\epsilon(\mathbf{m}^*)} \exp\left[\textsc{Ent}(\mathbf{v}) - \log|S|\right] \tag{22}$$

For any vector $\mathbf{v} \in \mathrm{Conv}(\mathbf{m}(S))$, Lemma 2 provides an upper bound for each term in the sum $\textsc{Ent}(\mathbf{v}) - \log|S| \leq -\sum_{i=1}^{k} \frac{(v_i - m_i^*)^2}{\max\{\tilde{m}_i^*, \tilde{v}_i\}} + 3k \log n$. We proceed by considering the least favorable vector $\mathbf{v}^* \in \mathcal{B}^c_\epsilon(\mathbf{m}^*)$ such that our lower bond on the decay of entropy (upper bound on probability) becomes smallest. Since, we are requiring coordinate-wise concentration, such point would differ from the optimal vector only in one-coordinate, and in particular in the coordinate that minimizes our lower bound. Any such vector $\mathbf{v}^* \in \mathcal{B}_\epsilon(\mathbf{m}^*)$, would have at least one coordinate $i \in [k]$ such that $|v_i - m_i^*| = \epsilon m_i^*$. Using the facts that that $\max\{\tilde{m}_i, \tilde{v}_i\} \leq \tilde{m}_i + \tilde{v}_i \leq (1+\epsilon) m_i^*$, we get

$$\textsc{Ent}(\mathbf{v}) - \log|S| \leq -\frac{\epsilon^2 (m_i^*)^2}{(1+\epsilon) m_i^*} + 3k \log n = -\frac{\epsilon^2}{(1+\epsilon)} m_i^* + 3k \log n \tag{23}$$

Now, by definition of the thickness $\mu(S) \leq \tilde{m}_i^*$ for all $i \in [k]$, and so a a vector $\mathbf{v}^*$ that minimizes the bound is such that $\textsc{Ent}(\mathbf{v}^*) - \log|S| \leq -\frac{\epsilon^2}{(1+\epsilon)}\mu(S) + 3k \log n$. Performing the union bound over $\mathcal{B}^c_\epsilon(\mathbf{m}^*)$ and utilizing Proposition 4, we get

$$\begin{aligned}
\mathbb{P}_S(\mathcal{B}^c_\epsilon(\mathbf{m}^*)) &\leq |\mathcal{B}^c_\epsilon(\mathbf{m}^*)| \cdot \mathbb{P}_S(\mathbf{v}^*) && (24)\\
&\leq \exp\left[-\frac{\epsilon^2}{(1+\epsilon)}\mu(S) + 5k \log n\right] && (25)\\
&\leq \exp\left[-\mu(S)\left(\frac{\epsilon^2}{1+\epsilon} - \frac{5k \log n}{\mu(S)}\right)\right] && (26)
\end{aligned}$$

Finally, identifying $\lambda(S)$ in the expression provides the statement. We note here that the resolution $r(S)$ is defined exactly so that the expression in the exponent is negative. The condition $\lambda(S) \ll 1$ is a requirement that makes concentration possible in a small scale, i.e $\epsilon \ll 1$. □

Tightness. The crucial steps in the proof are, firstly, the approximation of the log-partition function and, secondly, the L_2 distance bounds on the decay of entropy away from the optimum. Both steps are essentially optimal under general assumptions, as is shown in Proposition 2. Our proof can only improved by using higher order Stirling approximations and a more complicated integration process (incremental union bounds over L_∞-shells) instead of the simple union bound, to reduce the error at best from $5k \log n$ down to the minimum of $2k \log n$. Since, the above considerations would complicate the proof significantly and the gain is a small improvement in the constant we deem this unnecessary.

6. Proof of Theorem 2

We will now use our concentration theorem to prove that $\mathcal{P}$-symmetric sets are (ϵ, δ)-sandwichable. Before presenting the proof of the theorem we state two preliminary lemmas.

6.1. Conditional Independence

The first lemma is a simple calculation showing that indeed conditional on a specific edge-profile $\mathbf{v} \in \mathbf{m}(S)$, the uniform measure decomposes in a product of $G(n, m)$-like distributions, i.e., to sample a uniform element from $\{G \in S | \mathbf{m}(G) = \mathbf{v}\}$, one can select for all $i \in [k]$ a v_i-uniform subset from part P_i.

Proposition 5. *Consider for all $i \in [k]$ disjoint sets of edges $I_i, O_i \subset P_i$ and define the events $A_i = \{G \in \mathcal{G}_n : I_i \subset E(G)$ and $O_i \cap E(G) = \emptyset\}$. Conditional on the edge profile of G being $\mathbf{v}$, the events are independent, i.e. it holds that: $\mathbb{P}_S(A_1 \cap \ldots \cap A_k | \mathbf{v}) = \prod_{i=1}^k \mathbb{P}_S(A_i | \mathbf{v})$.*

Proof of Proposition 5. Since $G \sim U(S)$ the distribution of G is by definition uniform on S. This also means that it is uniform on the subset of graphs

having edge profile $\mathbf{m} \in \mathbb{N}^k$.

$$\begin{aligned}\mathbb{P}_S\left(A_1 \cap \ldots \cap A_k | \mathbf{v}\right) &= \frac{\mathbb{P}_S\left(A_1 \cap \ldots \cap A_k \cap \mathbf{m}(G) = \mathbf{v}\right)}{\mathbb{P}_S(\mathbf{m}(G) = \mathbf{v})} \\ &= \frac{|A_1 \cap \ldots \cap A_k \cap \mathbf{m}(G) = \mathbf{v}|}{|\mathbf{m}(G) = \mathbf{v}|}\mathbb{I}_{\mathbf{m}(S)}(\mathbf{v})\end{aligned}$$

where the first equality follows from Bayes rule and the second due to uniformity and the fact that our symmetry assumption implies that membership in S depends only on the edge-profile $\mathbf{m}$. Recall that each set A_i imposes the requirement that the edges in I_i are included in G and that the edges in O_i are not included in G. Having conditioned on $\mathbf{v}$, we know that exactly v_i edges from P_i are included in G and that we can satisfy event A_i by selecting any subset of $v_i - |I|_i$ edges out of $P_i \setminus (I_i \cup O_i)$. For convenience set $|P_i| = p_i$, $|I|_i = s_i$, $|O_i \cup I_i| = n_i$, and let C_ℓ^n denote the number of ℓ-combinations out of an n element set (binomial coefficient). The number of valid subsets of P_i is then given by $C_{v_i - s_i}^{p_i - n_i}$. As the constraints imposed are separable, we have:

$$\frac{|A_1 \cap \ldots \cap A_k \cap \mathbf{m}(G) = \mathbf{v}|}{|\mathbf{m}(G) = \mathbf{v}|} = \frac{\prod_{i=1}^k C_{v_i - n_i}^{p_i - r_i}}{|\mathbf{m}(G) = \mathbf{v}|} = \prod_{i=1}^k \frac{|A_i \cap \mathbf{m}(G) = \mathbf{v}|}{|\mathbf{m}(G) = \mathbf{v}|}$$

which gives the required identity by exploiting again uniformity of the probability measure. □

6.2. Basic Coupling Lemma

Consider a set of random variables $X_1, \ldots, X_k$ with laws $\nu_1, \ldots, \nu_k$. A coupling between a set of random variables is a (joint) probability distribution ν, such that $\mathbb{P}_\nu(X_i) = \mathbb{P}_{\nu_i}(X_i)$ for $i \in [k]$, i.e. the marginals of the random variables are consistent with their laws. Let $A = \{1, \ldots, N\}$ be a finite set with N elements. Further let X be a uniform subset of m elements out of A, denoted as $X \sim \mathrm{Samp}(m, A)$, and Z a subset of A where each element of A is included with the same probability p, denote as $Z \sim \mathrm{Flip}(p, A)$.

Lemma 3. *Given a set A with N elements and a number m, define $p^\pm(m) = \frac{m}{(1 \mp \delta)N}$. There exists a coupling ν such that for the random variables $X \sim \mathrm{Samp}(m, A)$ and $Z^\pm \sim \mathrm{Flip}(p^\pm, A)$ it holds*

$$\mathbb{P}_\nu\left(Z^- \subseteq X \subseteq Z^+\right) \geq 1 - 2\exp\left(\frac{\delta^2}{3(1+\delta)m}\right). \tag{27}$$

Proof. Let ν be the joint distribution of $U_1, \ldots, U_N$ i.i.d uniform in $[0,1]$ random variables, and $U_{(m)}$ denote the m-th smallest such random variable. Define $X(U) = \{i \in A : U_i \leq U_{(m)}\}$ and $Z^{\pm}(U) = \{i \in A : U_i \leq p^{\pm}\}$ to be subsets of A that depend on the random variables $U_1, \ldots, U_N$. By construction it is easy to see that $X(U)$ and $Z^{\pm}(U)$ have the correct marginal distribution. Further, due to uniformity, the following equivalence holds:

$$Z^- \subseteq X \subseteq Z^+ \Leftrightarrow |Z^-| \leq |X| \leq |Z^+|$$

To analyze the second event define the "bad" events:

$$\begin{aligned} B_- &= \{u \in [0,1]^N : \sum_{i=1}^{N} \mathbb{I}(u_i \leq p_-) > m\} \\ B_+ &= \{u \in [0,1]^N : \sum_{i=1}^{N} \mathbb{I}(u_i \leq p_+) < m\} \end{aligned}$$

Each event can be stated as the probability that the sum $X_{\pm}$ of n i.i.d Bernoulli $p_{\pm}$ random variables exceeds (smaller than) the expectation $np_{\pm}$. By employing standard Chernoff bounds, we get:

$$\begin{aligned} \mathbb{P}_\nu(B_-) &= \mathbb{P}_\nu(X_- > m) = \mathbb{P}_\nu(X > (1+\delta)np_-) \leq \exp(-\frac{\delta^2}{3(1+\delta)}m) \\ \mathbb{P}_\nu(B_+) &= \mathbb{P}_\nu(X_+ < m) = \mathbb{P}_\nu(X < (1-\delta)np_+) \leq \exp(-\frac{\delta^2}{2(1-\delta)}m) \end{aligned}$$

The proof is concluded through the use of union bound:

$$\mathbb{P}_\nu(B_- \cup B_+) \leq \mathbb{P}_\mu(B_-) + \mathbb{P}_\nu(B_+) \leq 2\exp\left(-\frac{\delta^2}{3(1+\delta)}m\right)$$

This concludes the lemma. □

Using, this simple lemma and Theorem 3, we prove the *sandwich theorem.*

6.3. Main proof.

Recall, that our aim is to prove that the uniform measure over the set S is (ϵ, δ)-sandwichable by some product measure $G(n, \mathbf{Q})$.

Proof of Theorem 2. Given a $\mathcal{P}$-symmetric convex set S, consider $\mathbf{m}^*(S)$ the optimal edge-profile and define the $n \times n$ matrix $\mathbf{Q}^*(S)$ as: $Q^*_{u,v} = \frac{m^*_i}{p_i}, \forall \{u,v\} \in P_i$ and $i \in [k]$. Further, define $q^*_i := \frac{m^*_i}{p_i}, \forall \in [k]$ to be used later. In order to prove the required statement, we need to construct a coupling between the random variables $G \sim U(S)$, $G^{\pm} \sim G(n, (1 \pm \epsilon)\mathbf{Q}^*)$. By separating edges according to the partition, we express the edge set of the graphs as $E(G) = E_1 \cup \ldots \cup E_k$ and $E(G^{\pm}) = E_1^{\pm} \cup \ldots \cup E_k^{\pm}$.

Let ν denote the joint probability distribution of $N+1$ i.i.d. uniform random variables $U_1, \ldots, U_{N+1}$ on $[0,1]$. As in the coupling lemma, we are going to use these random variables to jointly generate the random edge-sets of G^-, G, G^+. Using U_{N+1}, we can first generate the edge profile $\mathbf{v} \in \mathbf{m}(S)$ from its corresponding distribution. Then, conditional on the edge profile $\mathbf{v} \in \mathbb{N}^k$, the probability distribution of G factorizes in $G(n,m)$-like distributions for each block (Proposition 5). Lastly, we associate with each edge e a unique random variable U_e and construct a coupling for edges in each block separately.

In our notation, $E_i \sim \text{Samp}(v_i, P_i)$ while $E_i^{\pm} \sim \text{Flip}(q_i^{\pm}, P_i)$. Using Lemma 3, we construct a coupling for each $i \in [k]$ between the random variables E_i, E_i^+, E_i^- and bound the probability that the event $E_i^- \subseteq E_i \subseteq E_i^+$ does not hold. Using the union bound over the k parts, we then obtain an estimate of the probability that the property holds across blocks, always conditional on the edge-profile $\mathbf{v}$. The final step involves getting rid of the conditioning by invoking the concentration theorem.

Concretely, define B_i the event that the i-th block does not satisfy the property $E_i^- \subseteq E_i \subseteq E_i^+$ and $\mathcal{B}_\epsilon(\mathbf{m}^*)$ the set appearing in Theorem 3. We have that $\mathbb{P}_\nu(G^- \subseteq G \subseteq G^+) = 1 - \mathbb{P}_\nu(\cup B_i)$. Conditioning on the edge profile gives:

$$\begin{aligned}
\mathbb{P}_\nu(\cup B_i) &\leq \mathbb{P}_\nu(\mathcal{B}_\epsilon^c(\mathbf{m}^*)) + \sum_{\mathbf{v} \in \mathcal{B}_\epsilon(\mathbf{m}^*)} \mathbb{P}_\nu(\cup B_i | \mathbf{v}) \mathbb{P}_\nu(\mathbf{v}) \\
&\leq \mathbb{P}_\nu(\mathcal{B}_\epsilon^c(\mathbf{m}^*)) + \max_{\mathbf{v} \in \mathcal{B}_\epsilon(\mathbf{m}^*)} \mathbb{P}_\nu(\cup B_i | \mathbf{v}) \\
&\leq \mathbb{P}_\nu(\mathcal{B}_\epsilon^c(\mathbf{m}^*)) + \max_{\mathbf{v} \in \mathcal{B}_\epsilon(\mathbf{m}^*)} \left[\sum_{i=1}^{k} \mathbb{P}_\nu(B_i | \mathbf{v}) \right]
\end{aligned}$$

The first inequality holds by conditioning on the edge profile and bounding the probability of the bad events from above by 1 for all "bad" profiles (outside of the concentration set). The second inequality, is derived by upper

bounding the probability of the bad event by the most favorable such edge-profile and the last inequality follows from an application of the union bound. Applying Theorem 3 provides a bound on the first term and then invoking Lemma 3 allows us to obtain the following upper bound for $\mathbb{P}_\nu(\cup B_i)$

$$\exp\left[-\mu(S)\left(\frac{\epsilon^2}{1+\epsilon}-\lambda(S)\right)\right]+2\max_{\mathbf{v}\in\mathcal{B}_\epsilon(\mathbf{m}^*)}\left[\sum_{i=1}^{k}\exp\left(-\frac{\epsilon^2}{3(1+\epsilon)}v_i\right)\right]$$

Hence, we see that the upper bound is monotone in v_i for all $i \in [k]$. Additionally, we know that for all $\mathbf{v} \in \mathcal{B}_\epsilon(\mathbf{m}^*)$ it holds that $\mathbf{v} \geq (1-\epsilon)\mathbf{m}^*$. Further, by definition we have $\mathbf{m}^* \geq \mu(S)$. The upper bound for $\mathbb{P}_\nu(\cup B_i)$ now becomes:

$$\begin{aligned}
&\exp\left[-\mu(S)\left(\frac{\epsilon^2}{1+\epsilon}-\lambda(S)\right)\right]+2k\exp\left[-\frac{\epsilon^2(1-\epsilon)}{3(1+\epsilon)}\mu(S)\right]\\
\leq\ &\exp\left[-\mu(S)\left(\frac{\epsilon^2}{1+\epsilon}-\lambda(S)\right)\right]+\exp\left[-\mu(S)\left(\frac{\epsilon^2(1-\epsilon)}{3(1+\epsilon)}-\frac{\log(2k)}{\mu(S)}\right)\right]
\end{aligned}$$

Finally, using $\epsilon < 1/2$ and $\log(2k)/\mu(S) \leq \lambda(S)$ we arrive at the required conclusion. □

7. Beyond Monotone Events: Probabilities of Local Events

Our concentration theorem, beyond monotone events, also allows the calculation of the probabilities of *local events*, i.e., events that depend on a small (but not necessarily bounded) number of edges. We achieve this by relating the probability of an event under the uniform measure over the set S to its product approximation $G(n, \mathbf{Q}^*)$. This will be useful in computing expectations of functions that depend only on a few edges or are are sums of such functions, e.g., counting the number of subgraphs of some kind.

The extent and accuracy of our estimation capacity depends on the concentration of the number of edges from each part around its expectation. A global summary statistic for this is the thickness $\mu(S)$. For our purposes it will be easier to work with a related quantity that we define next.

Definition 5. *The* nominal resolution *of a $\mathcal{P}$-symmetric convex set S is*

$$\hat{\epsilon} = \hat{\epsilon}(S) := \sqrt{c^*\lambda(S)} = \sqrt{\frac{c^* 5k\log n}{\mu(S)}}\ ,$$

where $c^* = c^*(\lambda(S)) := 2(1+\lambda+\sqrt{\lambda^2+2\lambda}) > 2$ is such that $\hat{\epsilon}^2/(1+\hat{\epsilon}) = 2\lambda$.

The nominal resolution is the value of ϵ for which the upper bound given by Theorem 3 becomes independent of $\mu(S)$ and equal to $\exp(-5k \log n)$. Our goal is to provide conditions such that we can approximate the probability of local events within a multiplicative error proportional to $\hat{\epsilon}$. Our conditions on the set S ensure that this accuracy is non-trivial (Proposition 6).

Definition 6. *Given $\gamma > 0$ and a function $f(n) > 2$, we call a $\mathcal{P}$-symmetric convex set $S \subseteq \mathcal{G}_n$, $(f(n), \gamma)$-*well conditioned *if:*

$$\begin{aligned} \mu(S) &\geq c^* \left(9 \max\{\gamma, 1\} f(n)\right)^2 5k \log n \\ q_i^* &\leq \frac{\gamma}{1+\gamma}, \quad \forall i \in [k] \ . \end{aligned}$$

The upper bound on the q_i^* in Definition 6 is needed because, as we will see, we will derive tight multiplicative approximations to the probabilities of local events, so if an edge has probability close to 1 then any event that involves the non-presence of that edge would have probability close to 0 and that might be smaller than the accuracy that the concentration theorem provides, i.e., of the probability of the event that concentration around the optimizer doesn't happen. We next formalize the definition of local events.

For the rest of this section, as in Proposition 5, we consider for each $i \in [k]$ disjoint, potentially empty sets $I_i, O_i \subseteq P_i$ and events of the form

$$A_i = A_i(G) := \{G \in \mathcal{G}_n : I_i \subseteq E(G) \wedge O_i \cap E(G) = \emptyset\} \ ,$$

denoting $n_i := |I_i| + |O_i|$. Clearly, any event can be written as $\cap_{i=1}^k A_i$, for some events $A_1, \ldots, A_k$ as above.

Definition 7 (Local Events)**.** *In a product measure $\mathbf{Q}^*$, an event $A = \cap_{i=1}^k A_i$ is $(f(n), M)$-local if*

$$n_i \leq \frac{4 \log n}{-\log(q_i^*(1-q_i^*))} \quad \forall i \in [k] \qquad \textit{and} \qquad \sum_{i=1}^k n_i \leq M < f(n) \ .$$

Both conditions in Definition 7 can be checked easily given $\mathbf{Q}^*$ and A. The upper bound on n_i ensures that the probability of A is greater than the probability that concentration fails. We stress once more that the need for such a bound is due to our insistence on deriving multiplicative approximations of event probabilities. We are now ready to state our theorem.

Theorem 5. *Let S be any $(f(n),\gamma)$-*well conditioned *$\mathcal{P}$-symmetric convex set with entropy maximizer $\mathbf{Q}^* = \mathbf{Q}^*(S)$. For any $(f(n),M)$-local event A,*

$$\left|\frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{Q}^*}(A)} - 1\right| \leq \frac{M}{f(n)} \ .$$

Before proving Theorem 7 we state an immediate user-friendly corollary.

Corollary 1. *A function $h : \mathcal{G}_n \to \mathbb{R}$ is called $(f(n),M)$-local if the event $\{h(G) = a\}$ is an $(f(n),M)$-local event for all $a \in \mathbb{R}$. For any such function,*

$$|\mathbb{E}_S[h(G)] - \mathbb{E}_{\mathbf{Q}^*}[h(G)]| \leq \frac{M}{f(n)}\mathbb{E}_{\mathbf{Q}^*}[|h(G)|] \ .$$

The corollary comes with an explicitly computable estimate about the quality of the approximation and, by linearity of expectation, it can be used to get bounds for arbitrary sums of local functions. Furthermore, applying it to moments of (sums of) local functions yields estimates about the distribution of random variables. For instance we can approximate the moment generating function sufficiently close to 0.

A simple example is counting the number of triangles, $T(G)$, of a random graph $G \sim U(S)$. Given three distinct vertices $i,j,k \in [n]$, we may define the event $T_{ijk} := \{G \in \mathcal{G}_n | \{i,j\},\{j,k\},\{i,k\} \in E(G)\}$, containing all graphs having a triangle on $\{i,j,k\}$. It is easy to see that this is a local event and that the indicator of this event is a local random variable (function). Letting $T(G) = \frac{1}{3!}\sum_{i\neq j\neq k}\mathbb{I}_{T_{ijk}}(G)$ be the number of triangles in G, using Corollary 1 and observing that all terms in the sum are non-negative, we get that $\mathbb{E}_S[T(G)] \in \left[1 - \frac{3}{f(n)}, 1 + \frac{3}{f(n)}\right]\mathbb{E}_{\mathbf{Q}^*}[T(G)]$. In fact, by observing that $T(G)^\ell$ can be written as a sum of terms each of which is a product of at most ℓ different indicators, we see that we can also approximate the ℓ-th moment with a multiplicative accuracy of $3\ell/f(n)$. For instance, if we apply it on the second moment we can then use these estimates along with Chebyshev's inequality to bound the probability $\mathbb{P}_S(|T(G) - \mathbb{E}_{\mathbf{Q}^*}[T(G)]| > t)$.

Proof of Corollary 1. Let $I_+ = \{x \in \mathbb{R}^+ \mid \exists G \in \mathcal{G}_n, f(G) = x\}$, and similarly for I_i and $\mathbb{R}^-$. Under any measure μ over $\mathcal{G}_n$,

$$\mathbb{E}_\mu[h(G)] = \sum_{x\in I_+} x\mathbb{P}_\mu(h(G) = x) - \sum_{y\in I_-}|y|\mathbb{P}_\mu(h(G) = y) \ . \tag{28}$$

Applying (28) with $\mu = U(S)$ and with $\mu = G(n, \mathbf{Q}^*)$ we see that $|\mathbb{E}_S[h(G)] - \mathbb{E}_{\mathbf{Q}^*}[h(G)]|$ equals

$$
\begin{aligned}
&= \left| \sum_{x \in I_+} x \left[\mathbb{P}_S(h(G) = x) - \mathbb{P}_{\mathbf{Q}^*}(h(G) = x) \right] \right. \\
&\qquad \left. - \sum_{y \in I_-} |y| \left[\mathbb{P}_S(h(G) = x) - \mathbb{P}_{\mathbf{Q}^*}(h(G) = x) \right] \right| \\
&\leq \frac{M}{f(n)} \left[\sum_{x \in I_+} x \mathbb{P}_{\mathbf{Q}^*}(h(G) = x) + \sum_{y \in I_-} |y| \mathbb{P}_{\mathbf{Q}^*}(h(G) = y) \right] \\
&= \frac{M}{f(n)} \mathbb{E}[|h(x)|] \ ,
\end{aligned}
$$

where in the second step we used the triangle inequality and Theorem 5. □

7.1. Proof of Theorem 5

We first estimate the probability of A conditional on the edge profile being a specific vector in the concentration region $\mathcal{B}_{\hat{\epsilon}}(\mathbf{m}^*)$ and then invoking concentration to get rid of conditioning. For clarity of presentation we start by stating some technical facts, which we prove in the next subsection.

Proposition 6. *Under the conditions of Theorem 5, let $\hat{\epsilon}$ denote the nominal resolution of S. Then*

$$
\begin{aligned}
\hat{\epsilon} &\leq \frac{1}{9 \max\{\gamma, 1\} f(n)} && (29) \\
\mathbb{P}_{\mathbf{Q}^*}(A) &\geq \hat{\epsilon}^{-1} \exp(-5k \log n) && (30) \\
n_i &\leq \hat{\epsilon} m_i^* \ . && (31)
\end{aligned}
$$

Moreover, for $\mathbf{v} \in \mathcal{B}_{\hat{\epsilon}}(\mathbf{m}^*)$, i.e., in the concentration region, for all $i \in [k]$:

$$
(1 - 2\hat{\epsilon} \max\{\gamma, 1\})^{n_i} \leq \frac{\mathbb{P}_S(A_i|\mathbf{v})}{\mathbb{P}_{\mathbf{Q}^*}(A_i)} \leq \left(\frac{1 + \hat{\epsilon} \max\{\gamma, 1\}}{1 - \hat{\epsilon} \min\{\gamma, 1\}} \right)^{n_i} \ . \qquad (32)
$$

We now combine the conditional estimates for all classes and use the concentration theorem to get rid of conditioning.

Proof of Theorem 5. To prove the lemma we use again the strategy of conditioning on the edge-profile vector. Recall that $\mathcal{B}_{\hat{\epsilon}}(\mathbf{m}^*)$ denotes the "concentration set" and $\mathcal{B}^c_{\hat{\epsilon}}(\mathbf{m}^*)$ its complement. Trivially,

$$\mathbb{P}_S\left(\mathcal{B}_{\hat{\epsilon}}(\mathbf{m}^*)\right) \min_{\mathbf{v}\in\mathcal{B}_{\hat{\epsilon}}(\mathbf{m}^*)} \mathbb{P}_S(A|\mathbf{v}) \leq \mathbb{P}_S(A) \leq \mathbb{P}_S\left(\mathcal{B}^c_{\hat{\epsilon}}(\mathbf{m}^*)\right) + \max_{\mathbf{v}\in\mathcal{B}_{\hat{\epsilon}}(\mathbf{m}^*)} \mathbb{P}_S(A|\mathbf{v}) \ .$$

For our choice of $\hat{\epsilon}$, Theorem 3 implies that $\mathbb{P}_S\left(\mathcal{B}^c_{\hat{\epsilon}}(\mathbf{m}^*)\right) \leq \exp(-5k\log n)$. Also, since A is $(f(n), M)$-local, by (30),

$$\mathbb{P}_{\mathbf{Q}^*}(A) \geq \hat{\epsilon}^{-1}\exp(-5k\log n) > 0 \ .$$

Thus, diving by $\mathbb{P}_{\mathbf{Q}^*}(A)$ yields

$$\left(1-n^{-5k}\right) \min_{\mathbf{v}\in\mathcal{B}_{\hat{\epsilon}}(\mathbf{m}^*)} \frac{\mathbb{P}_S(A|\mathbf{v})}{\mathbb{P}_{\mathbf{Q}^*}(A)} \leq \frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{Q}^*}(A)} \leq \hat{\epsilon} + \max_{\mathbf{v}\in\mathcal{B}_{\hat{\epsilon}}}(\mathbf{m}^*)\frac{\mathbb{P}_S(A|\mathbf{v})}{\mathbb{P}_{\mathbf{Q}^*}(A)} \ . \qquad (33)$$

Clearly, $\mathbb{P}_{\mathbf{Q}^*}(A) = \prod_{i=1}^k \mathbb{P}_{\mathbf{Q}^*}(A_i)$, while $\mathbb{P}_S(A|\mathbf{v}) = \prod_{i=1}^k \mathbb{P}_S(A_i|\mathbf{v})$ by Proposition 5. If the total number of edges specifying A is $m = \sum n_i \leq M$, we therefore see that (32) implies

$$(1-2\hat{\epsilon}\max\{\gamma,1\})^m \leq \frac{\mathbb{P}_S(A|\mathbf{v})}{\mathbb{P}_{\mathbf{Q}^*}(A)} \leq \left(\frac{1+\hat{\epsilon}\max\{\gamma,1\}}{1-\hat{\epsilon}\min\{\gamma,1\}}\right)^m \ . \qquad (34)$$

To prove the theorem's lower bound from (33), (34) we use the inequality $(1-x)^y \geq 1-xy,\ \forall x<1$ and the fact that $n^{-5k} \leq \hat{\epsilon}$ for any $k \geq 1$ to obtain:

$$\frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{Q}^*}(A)} \geq \left(1-n^{-5k}\right)\left(1-2\hat{\epsilon}\max\{\gamma,1\}m\right) \geq 1-3\hat{\epsilon}\max\{\gamma,1\}m \ .$$

For the upper bound we first use the inequality $(1-x)^y \geq 1-xy,\ \forall x<1$ to simplify the fraction in the denominator. Then, using the inequalities $(1+x)^m \leq 1+2xm$ for $xm<1$ for the numerator and $1/(1-x) \leq 1+2x$ for $x<1/2$ for the denominator, we get

$$\begin{aligned}\frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{Q}^*}(A)} &\leq \hat{\epsilon} + \left(1+2m\hat{\epsilon}\max\{\gamma,1\}\right)\left(1+2\hat{\epsilon}\min\{\gamma,1\}m\right) \\ &\leq 1+9\hat{\epsilon}\max\{\gamma,1\}m \ .\end{aligned}$$

Putting everything together, we obtain that $\left|\frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{Q}^*}(A)}-1\right| \leq 9\hat{\epsilon}m\max\{\gamma,1\}$. Finally, $9\hat{\epsilon}\max\{\gamma,1\} \leq \frac{1}{f(n)}$ by Proposition 6 and $m \leq M$ by locality. □

7.2. Proof of Proposition 6

Equation (29) follows readily from S being $(f(n),\gamma)$-well conditioned.

To prove (30) we note that $\mathbb{P}_{\mathbf{Q}^*}(A) \geq \prod_{i=1}^{k}(q_i(1-q_i))^{n_i}$, since $0 \leq q_i^* \leq 1$. By the definition of locality, this product is, in turn, bounded from below by

$$\prod_{i=1}^{k} \exp\left(\frac{4\log n}{-\log(q_i(1-q_i))}\log(q_i(1-q_i))\right) \quad , \geq \frac{\exp(-5k\log n)}{\hat{\epsilon}}$$

since $\log(\hat{\epsilon}) \geq -\log n \geq -k\log n$ for all $k \geq 1$.

To prove (31) we note that

$$n_i \leq \frac{4\log n}{-\log(q_i(1-q_i))} \leq \frac{4\log n}{\log 4}$$

while $\hat{\epsilon}m_i^* \geq \sqrt{c^*\lambda\mu(S)^2} \geq c^*\, 5k\log n\, f(n)\, 9\max\{\gamma,1\} \geq 180\log n > \frac{4\log n}{\log(4)}$.

For our claim regarding the removal of conditioning, let $|I_i| = s_i$ (so $|O_i| = n_i - s_i$). We start by writing an exact expression for $\mathbb{P}_S(A_i|\mathbf{v})$, i.e.,

$$\mathbb{P}_S(A_i|\mathbf{v}) = \frac{\binom{p_i-n_i}{v_i-s_i}}{\binom{p_i}{v_i}} = \frac{(v_i)_{s_i}(p_i-v_i)_{n_i-s_i}}{(p_i)_{n_i}} \tag{35}$$

where $(x)_n$ denotes the descending factorial, for which we have the following easy bounds $(x-n)^n \leq x(x-1)\ldots(x-n+1) \leq x^n$.

We now obtain a lower bound on $\mathbb{P}_S(A_i|\mathbf{v})$ as follows. Recall that $q_i^* = m_i^*/p_i$. To get (37) we use that $\mathbf{v} \in \mathcal{B}_{\hat{\epsilon}}(\mathbf{m}^*)$, i.e. $|v_i - \mathbf{m}_i| \leq \hat{\epsilon}\tilde{m}_i \leq \hat{\epsilon}m_i^*$ and that $s_i \leq n_i \leq \hat{\epsilon}m_i^*$. Equation (38) follows trivially from (37), while for (39) we use the assumption $q_i^* \leq \gamma/(1+\gamma)$.

$$\begin{aligned}
\mathbb{P}_S(A_i|\mathbf{v}) &\geq \frac{(v_i-s_i)^{s_i}(p_i-v_i-n_i+s_i)^{n_i-s_i}}{p_i^{n_i}} && (36)\\
&\geq \left((1-2\hat{\epsilon})\frac{m_i^*}{p_i}\right)^{s_i}\left(1-(1+2\hat{\epsilon})\frac{m_i^*}{p_i}\right)^{n_i-s_i} && (37)\\
&\geq (1-2\hat{\epsilon})^{s_i}(q_i^*)^{s_i}(1-q_i^*)^{n_i-s_i}\left(1-\frac{q_i^*}{1-q_i^*}2\hat{\epsilon}\right)^{n_i-s_i} && (38)\\
&\geq (1-2\hat{\epsilon})^{s_i}\mathbb{P}_{\mathbf{Q}^*}(A_i)\,(1-2\hat{\epsilon}\gamma)^{n_i-s_i} \quad . && (39)
\end{aligned}$$

Proceeding analogously yields the upper bound

$$\mathbb{P}_S(A_i|\mathbf{v}) \leq (1+\hat{\epsilon})^{s_i}\mathbb{P}_{\mathbf{Q}^*}(A_i)(1+\hat{\epsilon}\gamma)^{n_i-s_i}\left((1-\hat{\epsilon})\frac{\gamma}{1+\gamma}\right)^{-n_i}$$

from which the claimed upper bound follows readily.

8. Applications

A common assumption throughout the paper is the existence of a partition of the edges. The partition expresses prior information that we have about the setting at hand and should be considered given. 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 the extent that the given partition is fine enough to express the desired property S. The typical pipeline is: (i) translate prior information in a partition of edges $\mathcal{P}$, (ii) express the set of interest S as a specification on the edge-profile $\mathbf{m}$, (iii) solve the entropy-optimization problem on $\mathrm{Conv}(\mathbf{m}(S))$ 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. We proceed with some examples.

Linear Programs on Graphs. Consider that each possible edge has multiple attributes, that can be categorical (type of relation) or operational (throughput, latency, 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: $Q^*_{uv}(S) = [1 + \exp(\boldsymbol{x}_\ell^\intercal \boldsymbol{\lambda})]^{-1}$. This is particularly useful as it opens the possibility of using Linear Programming to design networks with desired properties that are at the same time amenable to analysis through their product measure approximation. A simple instance of which is the following example.

Navigability. A concrete example of the applicability of our framework is found in the context of network navigability. Kleinberg [14, 15] 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 require-

ment that for any vertex in the graph, the probability of having an edge to a vertex at distance in the range $[2^{i-1}, 2^i)$ (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 scale i. In [16], 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, but 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 be accommodated we focus on the Stochastic Block Model. Consider a social network consisting of q groups of sizes $(\rho_1, \ldots, \rho_q)n$, where $\rho_i > 0$ for $i \in [q]$. As the partition of edges is naturally induced by the vertex partition, for simplicity we adopt a double indexing scheme: instead of the edge-profile vector $\mathbf{m} \in \mathbb{R}^{\binom{q+1}{2}}$ we use a symmetric edge-profile matrix $\mathbf{M} \in \mathbb{R}^{q \times q}$.

Imagine that we already have a graph property S that is symmetric with respect to this partition, e.g., any linear or convex constraint in terms of the elements of the matrix $\mathbf{M}$, and we would like to enforce the additional property that a specific group of vertices $g \in [q]$ 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 $\epsilon > 0$. Denoting by S_ϵ this restriction of S it is easy to see that S_ϵ is *not* symmetric with respect to the partition. Nevertheless, our results are still useful. Using a well known connection to Multitype Branching Process [17], the non-existence of a giant component in SBMs can be recast as a condition on the number of edges between each block. Concretely, given the edge-profile matrix $\mathbf{M}$ and for a given group $g \in [q]$, define the $(q-1) \times (q-1)$ matrix: $T(\mathbf{M})_{ij} := \frac{m_{ij}}{n^2 \rho_i}, \forall i, j \in [q] \setminus \{c\}$ 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 $\|T(\mathbf{M})\|_2 < 1$, then with high probability no giant component exists.

Based on our results, the original property S will be well-approximated by the product measure corresponding to some set of parameters of the SBM. If these parameters are such that $\|T(\mathbf{M})\|_2 < 1$ holds, then a *typical* element

of S will have the desired property. Note that since S_ϵ is not symmetric, a typical case result is the best we can hope for. If, on the other hand, this is not the case, we can exploit that the set $Q = \{\mathbf{M} : \ \|T(\mathbf{M})\|_2 < 1\}$ is convex to determine which the constraints defining S to change so that the entropic maximizer enters the set Q. Notably there may be multiple ways to achieve this, and convexity is key to being able to find a good modification of S to, probabilistically, achieve S_ϵ.

8.1. A concrete example: the 2D Torus.

Below is one of the simplest settings our machinery can handle. We consider $n = (2L+1)^2$ vertices embedded on a two dimensional torus, where each node $i \in [n]$ is associated with a vector $v^{(i)} \in \{0, \pm 1, \ldots, \pm L\}^2$, and where the distance d_{ij} between two vertices is defined as the ℓ_∞ toroidal distance

$$d_{ij} := \max_{\ell \in \{1,2\}} \min \left\{ \left| v_\ell^{(i)} - v_\ell^{(j)} \right|, \left| v_\ell^{(i)} - v_\ell^{(j)} + L \right| \bmod (L+1) \right\} \ . \tag{40}$$

Partition. By definition there are precisely L possible values for the distance between two distinct vertices. This naturally leads us to define a partition of edges $\mathcal{P} = (P_1, \ldots, P_k)$ into $k = L$ parts where an edge $\{i, j\} \in P_\ell$ iff $d_{ij} = \ell$. This a natural example where the partition is defined organically by our prior information about the geometry of the vertices. The number of edges in each part is $p_\ell := |P_\ell| = 4\ell n$.

Symmetric property. What are interesting properties that are invariant under the partition induced by the metric? In the spirit of Erdős and Rényi, where the number of edges is specified, we can specify an upper bound on the *total length* of the edges of the graph and seek the properties of a uniformly random graph from that set. Concretely, we define the set $S(B) := \left\{ G \in \mathcal{G}_n \left| \sum_{e \in E(G)} d_e \leq B \right. \right\}$.

Edge profile. For a graph $G \in \mathcal{G}_n$ we define its edge profile $\mathbf{m} \in \mathbb{N}^L$ as the vector with coordinates $m_\ell(G) := |\{e \in E(G) | d_e = \ell\}|$ for $\ell \in [L]$. Thus, in terms of the edge profile $S(B) = \left\{ G \in \mathcal{G}_n \left| \sum_{\ell=1}^{l} \ell \, m_\ell(G) \leq B \right. \right\}$, i.e., $S(B)$ contains the graphs whose edge profile satisfies $\sum_{\ell=1}^{L} \ell \, m_\ell \leq B$.

Entropy Optimization. For convenience we convert edge profiles to probability vectors $a_\ell = m_\ell/p_\ell$ so that the entropy optimization problem reads:

$$\begin{aligned} \max_{\mathbf{a}\in[0,1]^L} \quad & -\sum_{\ell=1}^{L} p_\ell \left[a_\ell \log a_\ell + (1-a_\ell)\log(1-a_\ell)\right] \\ \text{subject to} \quad & \sum_{\ell=1}^{L} p_\ell a_\ell \ell \le B \end{aligned}$$

Define the point $B_{1/2} := 2n\sum_{\ell=1}^{L}\ell^2 = \Theta(\frac{L^3}{3}n)$ above which the constraint becomes inactive. The solution to the optimization problem is

$$a_\ell^*(\beta) = \frac{1}{1+\exp(\beta\ell)} \ ,$$

where β is the unique solution to $\sum_\ell p_\ell \left[1+\exp(\beta\ell)\right]^{-1} = \min\{B, B_{1/2}\}$. In fact, since the function $f(\beta) := \sum_\ell p_\ell a_\ell^*(\beta)$ is strictly monotone for all $B < B_{1/2}$, we can equivalently specify $S(B)$ as $S_\beta := S(f^{-1}(\beta))$. For any $\beta \in \mathbb{R}_+$ we define the product measure $\mathbf{Q}^*(\beta)$ as:

$$Q_{ij}^* = \frac{1}{1+\exp(\beta d_{ij})} \ . \tag{41}$$

Sandwichability. To analyze S_β using the product measure, we need to verify the quantitative requirements of Theorem 2, which mainly involve the thichkness parameter $\mu(\beta) := \mu(S_\beta)$. For any $\beta \ge 0$ we have that $\tilde{\mu}_\ell^* = \mu_\ell^* = p_\ell a_\ell^* = \frac{4\ell n}{1+\exp(\beta\ell)}$. Further, $\mu(\beta) = \tilde{m}_L^* = \frac{4Ln}{1+\exp(\beta L)}$. If we want a condition number $\lambda(S_\beta)$ of the order of $1/\log n$ this requires $\mu(\beta) \ge 5L\log^2 n$ and, therefore,

$$\beta \le \beta_0 := \frac{\log(4n) - \log(5\log^2 n)}{L} \ . \tag{42}$$

For all such $\beta \in [0, \beta_0]$, Theorem 2 implies that the uniform measure on S_β is $\left(\sqrt{\frac{24}{\log n}}, 2n^{-5L}\right)$-*sandwichable.*

We can now answer any question we would like using the product measure $\mathbf{Q}^*$ rather than resorting to working with the rather inaccessible uniform measure on $S(B)$. To illustrate how our tools apply, we will show that with high probability the degree of every vertex is close to its expectation $\bar{d}(\beta) := \sum_{\ell=1}^{L} \frac{8\ell}{1+\exp(\beta\ell)}$.

Lemma 4. *For G picked uniformly at random from S_β with $\beta \le \frac{\log n}{2L}$,*

$$\mathbb{P}_{S_\beta}\left(\bigcup_{i\in[n]}\left(|d_i(G)-\bar{d}(\beta)|>2\epsilon\bar{d}\right)\right) \le 4\exp\left(-\frac{16}{\log n}\sqrt{n}+\log n\right) \ ,$$

where $\epsilon=\sqrt{24/\log n}$.

Proof. Let $A=\bigcup_{i\in[n]}(d_i-\bar{d}>2\epsilon\bar{d})$ (respectively $B=\bigcup_{i\in[n]}(d_i-\bar{d}<-2\epsilon\bar{d})$) be the event (set) that there is at least one vertex with degree a multiplicative constant larger (respectively smaller) than $\bar{d}$. To bound $\mathbb{P}_{S_\beta}(A\cup B)$ we will exploit the sandwich theorem and standard Chernoff and union bounds.

Set $\beta_0=\frac{\log n}{2L}$. For $0\le\beta\le\beta_0$ we see that $\bar{d}(\beta)\ge\bar{d}(\beta_0)=\sum_{\ell=1}^{L}\frac{8\ell}{1+n^{\frac{\ell}{2L}}}\ge 4\sqrt{n}$, since $2L=(\sqrt{n}-1)$ by definition. By the Chernoff bound, for any vertex $i\in[n]$, $\mathbb{P}_{Q^*}\left[d_i-\bar{d}>\epsilon\bar{d}\right]\le\exp\left(-\frac{\epsilon^2}{3}\bar{d}\right)$. The union bound over all n vertices gives $\mathbb{P}_{Q^*}\left[\bigcup_{i\in[n]}(d_i-\bar{d}>\epsilon\bar{d})\right]\le\exp\left(-\frac{\epsilon^2}{3}\bar{d}+\log n\right)$.

Since the set A is monotone increasing, the sandwich theorem yields

$$\mathbb{P}_{S_\beta}(A)\le\mathbb{P}_{(1+\epsilon)Q^*}(A)+2n^{-5L} \ .$$

Observing that $d_i-\bar{d}>2\epsilon\bar{d}\Leftrightarrow d_i-(1+\epsilon)\bar{d}>\frac{\epsilon}{1+\epsilon}(1+\epsilon)\bar{d}$ and applying Chernoff bounds for the product measure $(1+\epsilon)\mathbf{Q}^*$ gives

$$\begin{aligned}\mathbb{P}_{S_\beta}(A) &\le \exp\left(-\frac{\epsilon^2}{3(1+\epsilon)}\bar{d}+\log n\right)+2n^{-5L}\\ &\le 2\exp\left(-\frac{16}{\log n}\sqrt{n}+\log n\right) \ .\end{aligned}$$

Similarly, $d_i-\bar{d}<-2\epsilon\bar{d}\Leftrightarrow d_i-(1-\epsilon)\bar{d}<-\frac{\epsilon}{1-\epsilon}(1-\epsilon)\bar{d}$, yielding

$$\begin{aligned}\mathbb{P}_{S_\beta}(B) &\le \mathbb{P}_{(1-\epsilon)Q^*}(B)+2n^{-5L}\\ &\le \exp\left(-\frac{\epsilon^2}{2(1-\epsilon)}\bar{d}+\log n\right)+2n^{-5L}\\ &\le \exp\left(-\frac{48}{\log n}\sqrt{n}+\log n\right)+2n^{-5L}\\ &\le 2\exp\left(-\frac{48}{\log n}\sqrt{n}+\log n\right)\end{aligned}$$

Finally we have $\mathbb{P}_{S_\beta}(A\cup B)\le\mathbb{P}_{S_\beta}(A)+\mathbb{P}_{S_\beta}(B)\le 4\exp(-\frac{16}{\log n}\sqrt{n}+\log n)$. □

References

[1] P. Erdős, A. Rényi, On random graphs, Publicationes Mathematicae Debrecen 6 (1959) 290–297.

[2] E. N. Gilbert, Random graphs, Ann. Math. Statist. 30 (4) (1959) 1141–1144.

[3] B. Bollobás, Random graphs, 2nd Edition, Vol. 73 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2001.

[4] S. Janson, T. Łuczak, A. Rucinski, Random graphs, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[5] N. Linial, E. London, Y. Rabinovich, The geometry of graphs and some of its algorithmic applications, Combinatorica 15 (2) (1995) 215–245.

[6] A. Goldenberg, A. X. Zheng, S. E. Fienberg, E. M. Airoldi, A survey of statistical network models, Found. Trends Mach. Learn. 2 (2) (2010) 129–233.

[7] M. Newman, Networks: an introduction, Oxford University Press, 2010.

[8] F. Chung, L. Lu, Complex Graphs and Networks (CBMS Regional Conference Series in Mathematics), American Mathematical Society, Boston, MA, USA, 2006.

[9] R. Durrett, Random graph dynamics, Vol. 20, Cambridge university press, 2007.

[10] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, Z. Ghahramani, Kronecker graphs: An approach to modeling networks, The Journal of Machine Learning Research 11 (2010) 985–1042.

[11] L. Lovász, Large Networks and Graph Limits, Vol. 60 of Colloquium Publications, American Mathematical Society, 2012.

[12] A. Frieze, R. Kannan, Quick approximation to matrices and applications, Combinatorica 19 (2) (1999) 175–220.

[13] C. Borgs, J. T. Chayes, H. Cohn, Y. Zhao, An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions, arXiv preprint arXiv:1401.2906, 2014.

[14] J. M. Kleinberg, Navigation in a small world, Nature 406 (6798) (2000) 845.

[15] J. M. Kleinberg, Small-world phenomena and the dynamics of information, in: T. G. Dietterich, S. Becker, Z. Ghahramani (Eds.), NIPS, MIT Press, 2001, pp. 431–438.

[16] D. Achlioptas, P. Siminelakis, Navigability is a robust property, in: Algorithms and Models for the Web Graph, Springer, 2015, pp. 78–91.

[17] C. Mode, Multitype branching processes: theory and applications, Modern analytic and computational methods in science and mathematics, American Elsevier Pub. Co., 1971.