

Convex Random Graph Models

Dimitris Achlioptas^{*†}
Department of Computer Science
University of California, Santa Cruz
optas@cs.ucsc.edu

Paris Siminelakis[‡]
Department of Electrical Engineering
Stanford University
psimin@stanford.edu

Abstract

We propose a principled framework for designing random graph models from minimal assumptions. Our central principle is to study the uniform measure over symmetric subsets of graphs. The symmetries we require are far less limiting than those in existing graph models and flexible enough to encompass both geometric features and properties of graph limits. Our main contribution is to derive natural sufficient conditions under which the uniform measure over a set of graphs can be approximated by a product measure over its edges. In particular, we prove that often the uniform measure over the set of graphs with a highly complex, global graph property collapses asymptotically to a distribution in which edges appear independently with different probabilities, the probability of each computable from the property.

*A longer version of this work will be made available on the Arxiv.

†Research supported by a European Research Council (ERC) Starting Grant (StG-210743) and an Alfred P. Sloan Fellowship.

‡Supported in part by an Onassis Foundation Scholarship.

1 Introduction

Since their introduction in 1959 by Erdős and Rényi [23] and Gilbert [27], respectively, $G(n, m)$ and $G(n, p)$ random graphs have nearly monopolized the mathematical study of random networks [10, 34]. Given n vertices, $G(n, m)$ selects uniformly among all graphs with m edges, i.e., it includes a uniformly random m -subset of edges, whereas $G(n, p)$ includes each possible 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 [7, 10]. Due to their mathematical tractability these three models are a cornerstone of Probabilistic Combinatorics [4] and have found numerous applications in other domains such as the Analysis of Algorithms [1, 35], Coding Theory [55], Economics [32, 22], Game Theory [60, 18], and Statistical Physics [48].

This mathematical tractability stems from symmetry: the probability of each edge is either the same, as in $G(n, p)$ and $G(n, m)$, or merely a function of the overall potency of its endpoints, as in the configuration model. This extreme symmetry bestows numerous otherworldly properties to them, most prominent of which is near-optimal expansion. The homogeneity of the model can be interpreted as a lack of geometry, a dramatic manifestation of which is that the shortest path metric of such graphs suffers maximal distortion when embedded in Euclidean space [43]. 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 task. The difficulty lies in achieving a balance between realism and analytical 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 [29] and the vast majority of them only provide heuristic arguments to back up their claims. For a gentle introduction the reader is referred to the book of Newman [50] and for a more mathematical treatment to the books of Chung and Lu [17] and of Durrett [21]. Roughly speaking, random graph models can be organized into three broad categories: sequential generative models (e.g., preferential attachment, copying models, affiliation networks), geometric models (e.g., geometric random graphs, dot-product random graphs), and Erdős-Rényi variants (e.g., stochastic block models, Inhomogeneous Random Graphs).

Sequential generative models are based on the premise that there is an underlying evolutionary process that drives graph formation [20]. The archetypical such process is Preferential Attachment [54, 6], while other prominent examples are Copying models [37, 39] and models aiming for triadic closure [61, 41]. Although these models have been partly successful in explaining important phenomena, they employ a number of parametric, uniformity, and independence assumptions that make the attained results non-robust to even slight perturbations. Moreover, with few exceptions [8, 14], the very nature of the processes is only loosely motivated and each analysis is conducted in an ad hoc manner.

Random graphs influenced by geometry comprise, on the other hand, a general paradigm for proposing realistic models. The vanilla version [53] consists of fixing a geometry or a homogeneous spatial distribution for the vertices and connecting nodes within a certain distance. Whilst this model has important applications in wireless and sensor networks [31], its too homogeneous to be used as a general network model. One attempt to fix this deficiency is to repeat the Erdős-Rényi approach, i.e., to have conditionally independent edges where the probability of each edge is a function of the distance between its endpoints, while another is to incorporate some sequential growth process. Examples are the Watts-Strogatz Model [62], Dot-Product Random graphs [63, 52], Geometric Preferential Attachment [24, 3, 33] and Hyperbolic Random Graphs [38, 30]. These models can mimic first (e.g., degree distribution) and, in part, second order (e.g.,

clustering) properties of networks, but are still far [49, 58] from encompassing many other aspects, such as community structure [42] and degree correlations [51]. Overall, while technically tractable and more realistic, all too often this approach does not leave much to the imagination. In particular, the spatial homogeneity required for mathematical analysis, combined with the conditional independence of edges, means that typical properties can often be gleaned by considering the characteristics of a single vertex.

1.1 Related Work and Erdős-Rényi variants

The most general tractable model to date are Inhomogeneous Random Graphs, introduced by Bollobás, Janson, and Riordan [11]. The generality and utility of the model stems from allowing two levels of modeling, i.e., both on vertices and on edges. Specifically, each of the n vertices is assigned a feature vector x_i in some separable metric space \mathcal{S} and one takes a joint probability distribution μ on \mathcal{S}^n , such that the corresponding point process on \mathcal{S} is well-defined and has a limit. Then, conditional on the vector \mathbf{x}_n , edges are formed *independently*, with probability $p_{ij} = \min\{\kappa(x_i, x_j)/n, 1\}$, where κ is a kernel on $\mathcal{S} \times \mathcal{S}$. The authors provide very general results about many graph properties, most notably regarding the existence or not of a giant component, but also about the asymptotic degree distribution and diameter. Inhomogeneous random graphs form a general theoretical framework within which many different models can be unified. At the same time, though, it offers no insight on how to set either of the two key global components μ and κ in order to focus the measure on graphs with a desired set of properties.

A different approach was undertaken by Frieze, Vempala, and Vera [26] in introducing Log-concave random graphs. Such graphs are formed by first sampling a vector $\mathbf{x} = (x_1, \dots, x_{\binom{n}{2}})$ from a log-concave density F , and then considering the edge set $E_{F,p} = \{e : X_e \leq p\}$ for given $p > 0$. The authors provide many results about monotone properties like the existence of the giant component, connectivity and Hamiltonicity. The analysis of the model is enabled by expressing events of interest in terms of moments of individual coordinates (edges) of the distribution F . This indicates that biases on individual edges are only present through the marginals, and hence, due to the high-dimensionality of the setting, the sampling distribution F must be severely “biased” in order for interesting phenomena to appear. For instance, they propose to construct an explicit vector \mathbf{x}_H , such that for a given subgraph H the probability of H appearing is zero. The caveat is that this is achieved by *explicitly* setting the probability of each such possible subgraph to be zero. Arguably, one would like a model instead, that might not require such strict control on the micro-level of individual edges, but can still be used to generate graphs with prescribed properties.

The work most closely related to ours is the much heralded idea of Exponential Random Graphs [59, 56]. In this model, one seeks to define a probability distribution on the set \mathcal{G}_n of all graphs with n vertices such that the *expectation* of m functions $H_k(G)$ satisfy some *affine constraints*. Out of all valid distributions, the max-entropy distribution is selected, having the form $\mathbb{P}_\beta(G) = \frac{1}{Z(\beta)} \exp(\beta \cdot \mathbf{H}(G))$, where β is a parameter vector acting as an inverse temperature and $\mathbf{H}(G) = (H_1(G), \dots, H_m(G))$ is an (energy) function measuring the deviation from the different prescribed expectations. While this is a very general way of incorporating desired features *without making independence assumptions*, there are a number of issues. A first one, at least conceptually, is that constraints are imposed only in expectation. That means that there is no hope of having the distribution’s support confined on a particular set, except of course in the “zero temperature” limit $|\beta_i| \rightarrow \infty$. Moreover, in many natural sub-cases, it was proven that even sampling computationally, can take exponential time [9]. Lastly, in most cases where the model is interesting, it becomes analytically intractable as probabilities of events can be calculated only through simulation. All in all, although this approach is conceptually appealing, it has remained largely unsatisfactory as a general modeling tool, as in the absence of structural assumptions the measure rapidly becomes impenetrable.

1.2 Motivation

In trying to replicate real networks one approach is to pile on features, creating increasingly complicated models, in the hope of matching observed properties. Arguably, though, the point of any model is prediction. The reason we study random graphs is to understand what (unknown) graph properties are *typically implied* by other (known) graph properties. For example, when we study $G(n, m)$ we are asking “what properties are typically implied by the property of having m edges?” and we cast the answer in terms of properties that hold with high probability in a uniformly random element of the set of all graphs with m edges.

Our work departs from efforts to explicitly bake-in features, adopting instead an agnostic approach. Specifically, we propose a framework for designing random graph models from minimal assumptions via the following general paradigm: given a space of objects \mathcal{X} , e.g., graphs on n vertices, delineate subsets $S \subset \mathcal{X}$ of interest by imposing some set of constraints, and study the uniform measure on S . To enable the study of complex sets S , develop tools for approximating the uniform measure on S by product measures and use the latter as a proxy. Following this program, we propose a hierarchy of random graph models leading to a general and tractable class of models we call *Convex Random Graphs*. As this probably sounds rather abstract, consider the following thought experiment which germinated our work:

Given a collection of n points on the plane and a finite spool of wire, what can be said about the set of all graphs that can be built on the given points with the given wire? What does a uniformly random such graph look like? How does it change as a function of the available wire?

If all edges had unit length, the thought experiment would, essentially, recover $G(n, m)$ random graphs, where m is the total wire length. (More precisely, conditional on the number \tilde{m} of edges present the measure would be $G(n, \tilde{m})$, where \tilde{m} would typically be very close to m .) Because, though, edges have different lengths, the cap on the wire breaks the symmetry, removing the independence among them.

This thought experiment motivates our work as follows. Moving towards greater realism, at a minimum, requires moving towards models in which not all edges have the same probability. At the same time, both the act of explicitly imposing independence and asking the modeler to specify each edge probability defeats a large part of the predictive purpose. A natural way to address both of these issues is to give up on *a priori* independence between edges, take an agnostic stance with respect to everything not explicitly modeled, and study the induced uniform measure. This is *not* because there are no independences among edges (and more general events) in real networks. On the contrary. The point is that besides being far more intellectually honest, the discovery of independences is extremely useful in and of itself.

1.3 Outline of the paper

In Section 2, we briefly sketch our approach and state our results concerning Convex Random Graphs. We then discuss the utility of our model as part of a framework for deriving probability distribution on graphs with desired properties and present examples of its applicability (Section 3). In Sections 4–6 we present the derivation of our model and discuss the various assumptions made in comparison to other existing models. In Section 7 we present the proof of our main result (Theorem 1) and in Section 8 we establish rigorous connection of our model with a model with independent edges, by proving Theorems 25 and 3. Lastly, we discuss in greater extent the range of properties that are expressible under the assumptions of our model in Section 9.

2 Our Contribution

Our aim are models that can readily incorporate diverse prior knowledge about the context at hand, most notably geometry or class membership, while avoiding defining the probability distribution explicitly at the level of individual edges or forcing independence assumptions. In spite of such genericity, we would still like for the model to be amenable to rigorous analysis and to support precise mathematical statements.

Our central premise is to study the uniform measure over structured subsets of \mathcal{G}_n , the set of all edges on n vertices. We interpret structure as a class of symmetries with respect to an *arbitrary* partition \mathcal{P} of the set of all possible *edges*. Specifically, our symmetry requirement is that edges within each part are exchangeable (permutation invariant), and we call the related sets of graphs \mathcal{P} -symmetric. This reduces the complexity of, otherwise arbitrary sets, while allowing the encoding of prior-information such as geometry. We wish to emphasize that considering partitions of edges rather than of vertices is a key conceptual contribution of our work, as it is precisely what enables the expression of geometry. As we show next, a host of previously studied models obey (more restrictive forms of) our symmetry requirement and that, interestingly, a strictly more restrictive symmetry is also embedded in the recent theory of Graph Limits. We define a *Partially Symmetric Random Graph* (PSRG) as the uniform measure over an arbitrary \mathcal{P} -symmetric set S .

As we will see, \mathcal{P} -symmetric sets and the corresponding PSRG model possess a number of appealing mathematical properties. First, the assumed invariance under permutations of edges within parts implies that membership in the set depends only on the edge-profile $\mathbf{m}(G)$, the number of edges selected from each part, of a graph. Thus, conditioning on the edge-profile causes the uniform measure over the set to factorize in $G(n, m)$ -like distributions with edges in different parts becoming conditionally independent. As a result, a PSRG can be generated by first drawing an edge-profile and then selecting uniformly the resulting number of edges from each part. In other words, the study of PSRG over S is reduced to the study of the induced distribution of the set of the edge-profile vectors $\mathbf{m}(S)$.

Understanding this structure is no trivial matter as $\mathbf{m}(S)$ can be arbitrarily complex. Hence, some sort of regularity is required. Our approach is to consider \mathcal{P} -symmetric sets such that $\mathbf{m}(S)$ is *convex* and dub the uniform measure over any such set S a *Convex Random Graph* ($\text{CRG}_{\mathcal{P}}(S)$). To analyze convex random graphs we identify a subset of $\mathbf{m}(S)$ within which the uniform measure on S is concentrated. Our main result is the following concentration theorem.

Theorem 1 (Implicit). *Given a \mathcal{P} -symmetric set S , let $\mathbf{m}^* \in \mathbb{N}^k$ be the solution of an entropy-optimization problem over $\mathbf{m}(S)$ and $\lambda(S), r(S), \mu(S)$ be some explicitly defined functions of \mathbf{m}^* . If $\mu(S) > 5k \log n$, then for $\epsilon > r(S)$:*

$$\mathbb{P}_S (|\mathbf{m}(G) - \mathbf{m}^*| \leq \epsilon \tilde{\mathbf{m}}^*) \geq 1 - \exp \left[-\mu(S) \left(\frac{\epsilon^2}{1 + \epsilon} - \lambda(S) \right) \right] \quad (1)$$

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

The strength of the theorem depends on the geometry of $\mathbf{m}(S)$ which is captured by: the *thickness* $\mu(S)$ (minimum number of edges of a part at the optimum), the *condition number* $\lambda(S) = 5k \log n / \mu(S)$, and $r(S) \approx \lambda/2 + \sqrt{\lambda}$ the *resolution* of the set. Furthermore, for each set S we consider a corresponding *Edge Block Model* ($\text{EBM}_{\mathcal{P}}(\mathbf{q}^*)$) on the same partition \mathcal{P} , wherein edges are added independently, with the probabilities of edges, $\mathbf{q}^*(S)$, depending only on the part they belong. Drawing upon the concentration theorem and the equivalence of $G(n, m)$ with $G(n, p)$ conditional on the number of edges, we prove that a convex random graph is *sandwiched* between two edge-block models.

Theorem 2. Let $G^\pm \sim \text{EBM}_{\mathcal{P}}(\mathbf{q}^\pm)$ and $G \sim \text{CRG}_{\mathcal{P}}(S)$, where \mathbf{q}^\pm is defined by $q_i^\pm = \frac{1-\epsilon}{1\mp\epsilon} q_i^*$ for all $i \in [k]$, and $\mathbf{q}^*(S)$ is as before. For sets $S \subset \mathcal{G}_n$, satisfying the conditions of Theorem :

$$G^- \subset G \subset G^+ \quad (2)$$

with probability at least $1 - 2 \exp \left[-\mu(S) \left(\frac{\epsilon^2}{12} - \lambda(S) \right) \right]$ and $\epsilon < 1/2$.

The sandwich theorem provides a concise summary of the properties of Convex Random Graphs in the regime where concentration holds: *for most purposes the model is approximately equivalent to an edge block model where probabilities are given by the optimal entropic point*. In fact, we can make a related statement about the probability of “local” events (precisely defined later) on a convex random graph, thus providing a rather complete picture about their local properties (e.g. subgraph counts)

Theorem 3. Given a \mathcal{P} -symmetric convex set $S \subset \mathcal{G}_n$ with $\mu(S) \gg (5k \log n)^2$, define $\hat{\epsilon} \triangleq \sqrt{c^* \lambda(S)}$ where $c^* \approx 2$ is explicitly defined. For an $(\hat{\epsilon}, \gamma)$ -local event A involving at most $m \leq \frac{\sqrt[4]{\mu(S)}}{\sqrt{c^* 5k \log n}}$ edges, it holds:

$$\left| \frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{q}^*}(A)} - 1 \right| \leq \frac{8\gamma}{\sqrt[4]{\mu(S)}} \quad , \quad (3)$$

where $\mathbb{P}_{\mathbf{q}^*}(A)$ denotes the probability under $\text{EBM}_{\mathcal{P}}(\mathbf{q}^*)$.

Recap Our results on Convex Random Graphs have identified an interesting phenomenon: independence can *arise* from minimal symmetry and smoothness assumptions (convexity) at a level of specification far removed from individual edges. The concomitant analytical tractability promotes the use of Convex Random Graphs for designing random graph models with desired properties.

3 Convex Random Graphs as a Meta-model

Our investigation of Partially Symmetric and Convex Random Graphs was spurred by the desire to induce realism in random graph models in a principled, agnostic way. The language of edge partitions is used to express the specification of prior information while the concentration results obtained and the corresponding close connection with an Edge Block Model, under the convexity assumption for edge-profiles, are very powerful in terms of analytical tractability. Next we explore, in brief, the range of properties that are expressible in terms of edge-profiles for a given partition and outline the general paradigm we are proposing.

Probabilistic Reductions. At first glance, Partially Symmetric Random Graphs only allow the study of sets defined in terms of the edge-profile \mathbf{m} . While strictly true, this is grossly misleading. The reason being that while not many graph properties are *exactly* expressible in terms of \mathbf{m} , many interesting properties are *probabilistically reducible* to a specification over $\mathcal{M}_{\mathcal{P}}$, the set of edge-profiles. Specifically, the precise knowledge of the conditional distribution $\mathbb{P}(\cdot|\mathbf{m})$, arising from its factorization, motivates the following (\mathcal{P}, ϵ) -approximation of **any** (non-symmetric) set S :

$$\hat{S}_{\mathcal{P}}(\epsilon) \triangleq \{ \mathbf{m} \in \mathcal{M}_{\mathcal{P}} : \mathbb{P}(S|\mathbf{m}) \geq 1 - \epsilon \} \quad . \quad (4)$$

Observe that membership in $\hat{S}_{\mathcal{P}}(\epsilon)$ depends only on \mathbf{m} and, hence, the set is \mathcal{P} -symmetric. Of course, a priori, one is not guaranteed that $\hat{S}_{\mathcal{P}}(\epsilon)$ has a concise description. The extent to which this is true depends on the amount of information about S encoded in the partition \mathcal{P} . Nevertheless, there are two general avenues to obtain a succinct approximation:

1. **Phase Transitions:** a large part of random graph theory consists of establishing threshold functions for monotone properties. Such, results have been attained for Erdős–Rényi-like models and include properties such as the existence of a giant component, connectivity, the appearance of subgraphs etc. In cases where the partition at hand is congruent to those models, the set $\hat{S}_{\mathcal{P}}(\epsilon)$ is reduced to a simply inequality of the form $\{f_S(\mathbf{m}) \geq t(\epsilon)\}$. In such cases, the approximation is also *tight*, enabling the study of S itself.
2. **Sufficient Conditions:** in many settings, where we are not interesting in studying property S *per-se* but rather graphs *with* the property S (among other properties), sufficient conditions are acceptable. Such conditions can be extracted both *deterministically*, through graph theory, as well as *probabilistically*, by exploiting tractability of the conditional probability distribution.

Meta-model of Random Graphs. Imagine that one has information about vertices in a network, given by a set of *features* \mathbf{X}_n for vertices and potentially a set of features \mathbf{Y}_n for edges. The features can reflect membership in groups, geometric location, or even frequency of communication between two vertices. We are interested in incorporating this information and generate a random graph that, in addition, satisfies a set of properties S_1, \dots, S_ℓ . Observe now that, crucially, convex sets are closed under intersection, so that our results on Convex Random Graphs motivate the following general paradigm for deriving structured random graph models:

Algorithm 1 Convex Random Graph Approximation

Input: a collection of features $F_n = (\mathbf{X}_n, \mathbf{Y}_n)$, and properties $S_1, \dots, S_\ell \subset \mathcal{G}_n$.

- 1: *Partition:* using F_n obtain a partition \mathcal{P}_F of the $\binom{n}{2}$ possible edges between vertices.
- 2: *Symmetrization:* obtain \mathcal{P} -symmetric approximations $\hat{S}_1, \dots, \hat{S}_\ell$ of the properties.
- 3: *Convex Approximation:* approximate $\hat{S} = \hat{S}_1 \cap \dots \cap \hat{S}_\ell$ by a convex set \tilde{S} .
- 4: *Optimization:* Solve a max-entropy optimization problem to obtain $\mathbf{q}^*(\tilde{S})$, a vector of probabilities.

Output: the $(\mathcal{P}_F, \mathbf{q}^*)$ -Edge Block Model.

Note that the above approach could also be carried out by *starting* with an Edge Block Model and expressing the desired constraints in terms of the edge-probabilities. However, in such a case, the end result would be yet another arbitrarily constructed random graph model making strong independence *assumptions*, specifying the probability distribution at the level of *individual edges*. Our work can also be seen as a rigorous exploration of sufficient conditions for such proposals to be justified, as capturing the uniform measure on a set of interest. We next present two natural examples where the modeling capabilities of our framework are clearly manifest.

3.1 Application to Social Networks

Consider a social network with n vertices that are partitioned in ℓ communities with sizes $\rho_i n$ for $i \in [\ell]$. The vertex-partition induces a natural edge-partition \mathcal{P}_{SBM} , by assigning edges between different communities to different parts. Assume now that one wants to impose a number of constraints, e.g., that there is an overall density of edges $S_1 = \{\mathbf{1}^T \mathbf{m} \geq c \cdot n\}$ and that the number of triangles in a particular class i is small $S_2 = \left\{ \binom{n\rho_i}{3} \left(\frac{m_{ii}}{n^2 \rho_i^2} \right)^3 \leq t \right\}$. Observe, that in order to express the second property in the symmetry specifications, we considered an approximation of the number of triangles by its expectation under $\mathbb{P}(\cdot | \mathbf{m})$. Approximating

the number of subgraphs by its expectation is a choice that is justified only when concentration holds. Standard conditions for the latter, in case of Erdős-Rényi random graphs can be found in [34].

The last point illustrates that, in general, there are ways to construct reasonable and accurate \mathcal{P} -symmetric approximations of properties of interest. To make the point even more strongly, imagine that we now require additionally that out of the ℓ classes there is a specific class j that must act as the “connector”, meaning that we require that there should be *no giant component* in the network unless we use connections to nodes in j . Define the following $(\ell - 1) \times (\ell - 1)$ matrix:

$$(T(\mathbf{m}))_{ij} \triangleq \frac{m_{ij}}{n^2 \rho_i}, \quad \forall i, j \in [\ell] \setminus j \quad (5)$$

which is a function of \mathbf{m} . A theorem of Bollobas, Janson and Riordan [11] implies:

Theorem 4 (Informal). *Let $\|\cdot\|_2$ denote the operator norm (maximum singular value). If $\|T(\mathbf{m})\|_2 \leq 1$ no giant component exists, while if $\|T(\mathbf{m})\|_2 > 1$ a giant component exists.*

Thus, in our framework, the property $S_4 = \{ \text{no giant component without vertices from } j \}$, not only can be accurately approximated under the partition \mathcal{P}_{SBM} by

$$\hat{S}_3 = \{ \mathbf{m} \in \mathcal{M}_{\mathcal{P}_{SBM}} : \|T(\mathbf{m})\|_2 \leq 1 \} \quad (6)$$

but \hat{S}_3 happens to also be convex! (See e.g., [16].) Since, properties S_1, S_2 are also convex, $S_1 \cap S_2 \cap S_3$ is convex and, therefore, we can bring to bear the full thrust of our concentration and sandwiching yielding an eminently tractable model for the uniform measure on graphs that satisfy these properties. In other words: convexity makes engineering complex objects possible.

3.2 Probabilistic Network Design via Linear Programs

To exemplify that highly intricate properties can be treated within our framework we next discuss the treatment of network navigability [36]. In [2], we prove that for *coherent* geometries, an abstraction of (the geometric properties of) grids and set-systems, a sufficient condition for a network to be navigable is to add $\Theta(\frac{n}{\log^p(n)})$ random edges at each distance-scale $\{\log \log n, \dots, \log n\}$, where p controls the power in the polylogarithmic-time routing. This fact, thus, permits us to consider navigability as yet another convex constraint in the context scale-induced partitions (defined in Section 4), namely :

$$\hat{S}_{\mathcal{P}}(\epsilon) = \{ \mathbf{m} \in \mathcal{M}_{\mathcal{P}} : m_i \geq \frac{n}{\log^p(n)} \} . \quad (7)$$

We can thus combine navigability with arbitrary linear constraints on the edge-profile. Specifically, form the matrix $A = [A_1 \ \dots \ A_k]$ and define the following convex \mathcal{P} -symmetric set:

$$S = \{ \mathbf{m} \in \mathcal{M}_{\mathcal{P}} : A \cdot \mathbf{m} \leq \mathbf{b} \in \mathbb{R}^\ell \} . \quad (8)$$

The matrix A can, for instance, encode i) navigability $S_1 = \{ I_n \cdot \mathbf{m} \geq \frac{Cn}{\log^p n} \mathbf{1} \}$, ii) bounded total cost $S_2 = \{ \mathbf{c}^T \mathbf{m} \leq B \}$, or any other linear constraint imaginable. That is, we can enforce desirable properties in a probabilistic sense in terms of an entire *linear program* (compare that to simply specifying the exact number of edges from each part...) And for sets defined by linear constraints, the entropy-optimization problem is not only tractable but has a closed form solution in terms of the dual variables $\lambda \in \mathbb{R}_+^\ell$:

$$q_i^*(S) = \frac{1}{1 + \exp(A_i^T \lambda)} \quad (9)$$

4 Structure via Symmetry

4.1 Boolean Random Graphs

Let \mathcal{G}_n be the set of all graphs on n vertices and let $H_n = \{0, 1\}^{\binom{n}{2}}$. A maximally agnostic approach to studying random graphs is to consider the uniform measure over arbitrary subsets of graphs.

Definition 5. For S an arbitrary non-empty subset of \mathcal{G}_n , the random variable G with uniform distribution over S is called a Boolean Random Graph and is denoted by $G \sim \mathcal{B}(S)$.

The term Boolean is motivated by the fact that the probability distribution $\mathbb{P}_S(G) = \frac{1}{|S|} \mathbb{I}_S(G)$, $\forall G \in \mathcal{G}_n$ where $\mathbb{I}_S(G)$ is the characteristic function of the set S , providing a connection between random graph models and Boolean Analysis. In this setting, we can formulate every imaginable set of properties, but this does not come cheap. Even just the description length of the function (set), either explicit by enumerating elements, or implicit through Fourier representation, can be of exponential size, rendering the model intractable in its full generality. Thus, to make progress, the first step is identifying a set of assumptions that endow \mathbb{I}_S with some structure, that is specify a “language” in which to express S .

4.2 Invariance under Permutations

Perhaps the most natural way of introducing structure is by imposing symmetry. This is formally expressed as the *invariance* of the function \mathbb{I}_S under the action of a group of transformations T , i.e., requiring that $\forall t \in T : \mathbb{I}_S(t(x)) = \mathbb{I}_S(x), \forall x$. Our approach to symmetry is to consider the permutation (symmetric) group on edges. In particular, we consider functions (sets) \mathbb{I}_S for which there is a partition \mathcal{P} of the $\binom{n}{2}$ edges such that the function is invariant under any permutation of the edges (indices) within a part of \mathcal{P} . Therefore, if graphs G, G' contain the same number of edges from each part, then $\mathbb{I}_S(G) = \mathbb{I}_S(G')$.

Definition 6. Let $\mathcal{P} = (P_1, \dots, P_k)$ be a partition of the set of all edges on n vertices and \mathcal{S}_ℓ be the set of permutations on ℓ indices. For the given partition and each part $i \in [k]$, let $x_{P_i} = (x_{j_1}, \dots, x_{j_{|P_i|}})$ denote the substring of $x \in H_n$ consisting of indices in P_i . Consider the set

$$\Pi_n(\mathcal{P}) = \left\{ \pi \in \mathcal{S}_{\binom{n}{2}} : \forall x \in H_n, \pi(x) = (\pi_1(x_{P_1}), \dots, \pi_k(x_{P_k})) \text{ and } \pi_i \in \mathcal{S}_{|P_i|}, \forall i \in [k] \right\}$$

of permutations acting within parts of the partition \mathcal{P} . A function that is invariant under the action of $\Pi_n(\mathcal{P})$ is called \mathcal{P} -symmetric or in general partially symmetric.

We argue that this notion of symmetry is both natural and well motivated by a number of facts. For one, considering a subgroup of the symmetric group is natural due to the Cayley Theorem [5] asserting that every group is isomorphic to a subgroup of the permutation group. More specifically, far more restrictive forms of the symmetry we require are in fact embedded in nearly all existing graph models.

Erdős-Rényi . If \mathcal{P} is the trivial partition consisting of a single part, then $\Pi_n(\mathcal{P}) = \mathcal{S}_{\binom{n}{2}}$, i.e., all edges are equivalent. Thus, $\mathbb{I}_S(G)$ depends only on the *number* of edges in G , so that the uniform measure over S gives rise, for example, to $G(n, m)$. Every probability distribution where edges are exchangeable obeys this symmetry, e.g., $G(n, p)$. This is the maximum amount of symmetry one can impose.

Boolean Random Graphs. At the other extreme, if \mathcal{P} consists of $\binom{n}{2}$ atomic parts, then $\Pi_n(\mathcal{P}) = I_n$ and there is no symmetry on the underlying set, recovering the Boolean Random Graphs of full generality.

Stochastic Kronecker Graphs. A more realistic model are Kronecker Graphs [40, 45, 28]. This model takes the k -fold Kronecker product of a 2×2 symmetric matrix $P_{[1]}$ of probabilities to form an $n \times n$ matrix $P_{[k]}$ where $n = 2^k$. Edges are included independently with probability $P_{[k]}(i, j) = \prod_{\ell=1}^k P_{[1]}(i_\ell, j_\ell)$ that depends on the binary representation of the labels of the two vertices. Consider the partition, $\mathcal{P}_{\text{Kron}}$ formed by assigning edges with the same probability in the same part. This partition has at most $3^k \approx n^{1.58}$ parts and the group $\Pi_n(\mathcal{P}_{\text{Kron}})$ consists of all permutations of edges with the same probabilities.

Stochastic Block Models. Another important family of random graph models are the so called Stochastic Block Models (SBM). These models presuppose the existence of a *vertex* partition $\mathcal{V} = (V_1, \dots, V_\ell)$ into ℓ blocks and the existence of a matrix of probabilities $P \in [0, 1]^{\ell \times \ell}$. A random digraph according to $SBM(\mathcal{V}, P)$ is produced by including every arc (i, j) independently with probability $p_{\bar{i}, \bar{j}}$, where \bar{i}, \bar{j} are the blocks of the vertex-partition containing i, j , respectively. (To generate undirected graphs, P is usually taken to be a symmetric matrix and an arc direction is ignored.) Clearly, the vertex partition \mathcal{V} induces an edge partition $\mathcal{P}_{\mathcal{V}} \triangleq \{V_i \times V_j\}_{i, j \leq \ell}$. Here the set $\Pi_n(\mathcal{P}_{\mathcal{V}})$ includes all permutations of edges that result from permutation of vertices within a block, thus falling under our general definition of symmetry. This is particularly important both due to the utility of SBM in theoretical investigations (e.g., [47, 57, 46]), and to exemplify that community structure can be encoded in our framework.

Relation to Graph Limits The fact that our symmetry requirement encompasses Stochastic Block Models is particularly pertinent in light of the theory of Graph Limits [44, 12]. According to that theory, given a sequence of graphs obeying some regularity conditions, one can extract a limiting object that captures many properties of the sequence as a measurable integrable¹ function $W : [0, 1]^2 \rightarrow [0, 1]$, called a *graphon*, see [13, 15]. Inherent in the construction of the limiting object is an intermediate approximation by a sequence of Stochastic Block Models obtained after invoking the (weak) Szemerédi Regularity Lemma [25, 12]. The conclusions we can draw for our purposes is that any property that is encoded in the limiting object is expressible within our framework of symmetry and particularly can capture densities of subgraphs of size depending on the regularity of the sequence [12, Section 2.9].

4.3 Enabling the Expression of Geometry

In all the examples of symmetry presented in the previous section, a crucial point was that the edge-partition had block-vertex symmetry, i.e., it *factorized* along vertices: there existed a partition of the vertices such that the part of an edge was a function of the parts of its endpoints. In our symmetry requirements we will **not** make such an assumption and this is perhaps the *most significant* feature of our model as it enables the expression of geometry. For example, we can partition edges according to “length”, so that each part contains edges of (approximately) equal length. As a result, our notion of symmetry is neither confined in models with a mean-field type structure, nor does it impose a specific geometry on the vertices, or an independence requirement between edges.

We define below the following notion of an *abstract geometric partition*.

Definition 7. Consider a symmetric non-negative valued function, e.g., distance, similarity, membership, $g : V \times V \rightarrow \mathbb{R}^+$ such that $g(i, j)$ denotes the “distance” between i and j . If $\mathcal{R} = (R_1, \dots, R_k)$ is a partition of \mathbb{R}^+ , the induced partition $\mathcal{P}_{g, \mathcal{R}} = \{P_1, \dots, P_k\}$, where $P_\ell \triangleq \{(i, j) \in V \times V : g(i, j) \in R_\ell\}$, is called an *abstract geometric partition based on g and \mathcal{R} or a (g, \mathcal{R}) -partition*.

¹In general L^p -integrable for some $p \geq 1$.

We can use such partitions both to unify previous models and to cast new models into our framework.

Geometric Random Graphs. The simplest variant of this model, assumes a probability distribution (usually uniform) on some geometric space \mathcal{X} , typically $[0, 1]^d$ for small d , and assigns randomly to each vertex i a point $X_i \in \mathcal{X}$. Points are then connected if and only if the distance $d_{\mathcal{X}}(i, j)$ is smaller than a threshold $r(n)$. If we consider the partition $\mathcal{R}_r = \{[0, r), [r, \infty)\}$, it is clear that the model is $(d_{\mathcal{X}}, \mathcal{R}_r)$ -symmetric. However, if one looks at the action of the corresponding symmetry group $\Pi_n((d_{\mathcal{X}}, \mathcal{R}_r))$ one finds that it only maps elements onto themselves. This is a consequence of the fact that given the location of the points, encoded in the function $d_{\mathcal{X}}$, the graph is deterministic.

This problematic aspect is generally addressed by considering that edges are added independently, with the probability of each edge being a function of the distance between its endpoints. The symmetry assumption then boils down to assigning equal probability to edges for which $d_{\mathcal{X}}(i, j)$ is in the same part of \mathcal{R} . Besides Geometric Random Graphs, many models, or slight variations thereof, can be shown to enjoy such symmetry, including Random Intersection Graphs [19] or Dot-Product Random Graphs [63, 52].

Generic Scale-Induced Partitions. Given a collection of vertices endowed with a distance function g , i.e., taking the geometry as *input*, we can scale g so that its minimum value is 1, set $\lambda = \lambda(\epsilon) = 1 + \epsilon > 1$ and take $R_i = [\lambda^{i-1}, \lambda^i)$. The partition $\mathcal{P}_\lambda = \{R_1, \dots, R_k\}$ where $k = \max\lceil \log_\lambda \text{diameter}(\mathcal{X}) \rceil$ is called a λ -scale partition and will be utilized in a latter discussion on Navigability in random graphs. Note that the edges in each part will all have the same value of g up to a factor of λ , i.e., we can have arbitrarily good accuracy, while for any fixed $\epsilon > 0$ and $\text{diameter}(\mathcal{X}) = \text{poly}(n)$ we end up with only $k = O(\log n)$ parts. (As we will see, we will derive strong results even when $k = \text{poly}(n)$, but it reassuring to see that even far fewer classes would be expressive in applications, e.g., if the vertices lied on any finite-dimensional lattice.)

5 Partially Symmetric Random Graphs

As we saw in the previous section, partial symmetry fruitfully introduces structure while leaving space for interesting limiting and geometric properties to be encoded. This motivates the following definition.

Definition 8 (Partially Symmetric Random Graphs). *Given a partition \mathcal{P} and \mathcal{P} -symmetric set S , the random variable $G \in \mathcal{G}_n$ with uniform distribution over S is called an (\mathcal{P}, S) -random graph. We denote that fact by $G \sim \mathcal{P}(S)$. Any such model is called a Partially Symmetric random graph.*

Combining the uniform measure with partial symmetry has a number of appealing consequences.

Representation. Since the characteristic function $\mathbb{I}_S(x)$ is invariant under $\Pi_n(\mathcal{P})$, we see that it depends only on the vector $\mathbf{m}(x) = (m_1(x), \dots, m_k(x))$ of the number of edges (Hamming weight) selected from each part. That is, $\mathbb{I}_S(x) = \mathbb{I}_{\mathbf{m}(S)}(\mathbf{m}(x))$, $\forall x \in H_n$, where $\mathbf{m}(x) \in \mathbb{R}_+^k$ and $\mathbf{m}(S) \subset \mathbb{R}_+^k$ denotes the image of S under \mathbf{m} . From here on, we will refer to $\mathbf{m}(x)$ (equivalently of $\mathbf{m}(G)$) as the *edge-profile*.

Proposition 9. *Given a partition \mathcal{P} , any function $f : H_n \rightarrow \mathbb{R}$ invariant under $\Pi_n(\mathcal{P})$, depends on $\mathbf{x} \in H_n$ only through the edge-profile $\mathbf{m}(\mathbf{x})$.*

Generation. By Proposition 9, one can generate an (\mathcal{P}, S) -random graph G via a two step process:

1. Sample an edge profile \mathbf{m} out of the distribution on \mathbb{N}_+^k induced by the uniform distribution on S .
2. Given $\mathbf{m} = (m_1, \dots, m_k)$ select a uniformly random m_i -subset of edges from each part P_i of \mathcal{P} and form G as the union of these subsets.

Conditional Independence. The decomposition of the generation process into two stages highlights a prominent feature of Partially Symmetric random graphs, namely conditional independence. Specifically, that conditioning on the edge profile renders edges in different parts independent.

Proposition 10. *Let $G \sim \mathcal{P}(S)$, for all $i \in [k]$ consider the disjoint sets of edges $I_i, O_i \subset P_i$ and define the events $A_i = \{G \in \mathcal{G}_n : I_i \subset E(G) \text{ and } O_i \cap E(G) = \emptyset\}$. Conditional on the edge profile \mathbf{m} of G being \mathbf{v} , the events are independent, i.e. it holds that:*

$$\mathbb{P}_S(A_1 \cap \dots \cap A_k | \mathbf{v}) = \prod_{i=1}^k \mathbb{P}_S(A_i | v_i)$$

moreover $\mathbb{P}_S(\cdot | \mathbf{v}) = \mathbb{P}(\cdot | \mathbf{v})$ depends on S only through whether $\mathbf{v} \in \mathbf{m}(S)$.

The proof of Proposition 10, while elementary, is instructive of the mechanics of Partially Symmetric random graphs. Given the edge profile \mathbf{m} , not only is the distribution of edges known but it factorizes to a product of $G(n, m)$ -like distributions for each part. Thus, the complexity of the uniform measure on S is manifested *entirely* in the induced distribution on $\mathbf{m}(S) \in \mathbb{N}^k$, whose structure we discuss next.

Definition 11. *Given an edge profile $\mathbf{m} \in \mathbb{N}^k$ define the entropy of \mathbf{m} as*

$$\text{ENT}(\mathbf{m}) = \sum_{i=1}^k \log \binom{p_i}{m_i}, \quad (10)$$

where $p_i = |P_i|$ denotes the number of edges in part i of partition \mathcal{P} .

Thus, S is decomposed into equivalence classes according to \mathbf{v} , with the size of each class given by its entropy (10). Using this decomposition, we immediately see that for all valid edge-profiles $\mathbf{v} \in \mathbf{m}(S)$,

$$\mathbb{P}_S(\mathbf{v}) = \frac{1}{Z(S)} e^{\text{ENT}(\mathbf{v})} \quad (11)$$

$$\mathbb{P}_S(G | \mathbf{v}) = \frac{1}{\exp(\text{ENT}(\mathbf{v}))} \mathbb{I}_{\mathbf{v}}(\mathbf{m}(G)) \quad (12)$$

where $Z(S) = \sum_{\mathbf{v} \in \mathbf{m}(S)} \exp(\text{ENT}(\mathbf{v})) = |S|$ is the normalizing constant (partition function).

5.1 The Entropy of Edge-Profiles

As we saw, understanding the uniform measure on S naturally reduces to understanding the distribution on the set of edge-profiles $\mathbf{m}(S)$ induced by the entropy function. This is where the real work begins and what makes the effort hard is that $\mathbf{m}(S)$ can be an *arbitrary* subset of $\mathcal{M}_{\mathcal{P}} \triangleq \prod_{i=1}^k \{0, \dots, p_i\}$. Generically, we would like to identify a subset $M \subset \mathcal{M}_{\mathcal{P}}$, a distance $d : \mathbb{R}^k \times \mathbb{R}^k$ and a radius $\Delta(n)$ such that

$$\mathbb{P}_S(d(\mathbf{v}, M) \geq \Delta) \rightarrow 0, \quad (13)$$

where $d(\mathbf{v}, A) = \inf_{\mathbf{x} \in A} d(\mathbf{v}, \mathbf{x})$. Concretely, a most natural approach towards this goal is to focus on the approximate modes of the distribution on $\mathbf{m}(S)$. For $\epsilon = \epsilon(n) > 0$ not necessarily small, let

$$M_{\mathcal{P}}^{\epsilon}(S) = \{\mathbf{v} \in \mathbf{m}(S) : \text{ENT}(\mathbf{v}) \geq \sup_{\mathbf{z} \in \mathbf{m}(S)} \text{ENT}(\mathbf{z}) - \epsilon\} , \quad (14)$$

where decreasing ϵ makes $M_{\mathcal{P}}^{\epsilon}(S)$ “simpler”, while increasing it improves concentration (Δ can be smaller).

In increasing order of value (and aspiration) the motivation for conducting such analysis is:

1. **Structure:** the set $M_{\mathcal{P}}^{\epsilon}(S)$ is both much smaller than $\mathbf{m}(S)$ and structured, as points are approximate maximizers of the entropy over $\mathbf{m}(S)$. For instance, one can try to get a handle on the set $M_{\mathcal{P}}^{\epsilon}(S)$ by attacking the corresponding optimization problem analytically but *also computationally*.
2. **Orthogonality:** the conditional decomposition of the uniform measure in a product of $G(n, m)$ -like distributions implies that the distance in (13) can account each of the k coordinates independently.
3. **Tractability:** ideally, the set $M_{\mathcal{P}}^{\epsilon}(S)$ can be shrunk to a point (by taking $\epsilon = 0$), implying localization around a single point \mathbf{m}^* . Combined with orthogonality, this implies that the uniform measure on S essentially reduces to $\mathbb{P}_S(G|\mathbf{m}^*)$, a completely tractable distribution.

Recap. Partially Symmetric random graphs satisfy many of the requirements we initially set. They were derived from the first principles of uniformity and permutation symmetry, both reflecting maximal agnosticism beyond what is explicitly considered known, and once we peel off the outer layer of complexity (conditioning on the edge-profile), they become tractable as the union of $G(n, m)$ -like subgraphs. At the same time, they present a trade-off between modeling power, with the number of parts in the partition of \mathcal{P} on one hand and the complexity of the geometry of $M_{\mathcal{P}}^{\epsilon}(S)$ on the other. In the next sections we establish the existence of an extremely fertile middle-ground, *Convex Random Graphs*. These allow us to banish the complexity of $M_{\mathcal{P}}^{\epsilon}(S)$ completely, achieving tractability, while accommodating a number of classes that grows with n at a rate reflecting the alignment between S and \mathcal{P} .

6 Convex Random Graphs

Our approach to tractability is to provide conditions under which the geometry of $M_{\mathcal{P}}^{\epsilon}(S)$ is simplified and, consequently, rigorous statements enabled. As a first step, we proceed with a re-parametrization of the edge-profile $\mathbf{m} \in \mathbb{N}^k$ in terms of effective probabilities $\mathbf{a} \in [0, 1]^k$, defined through the relation $a_i(G) \triangleq \frac{m_i(G)}{p_i}$ for all $i \in [k]$. The vector $\mathbf{a}(G) = (a_1(G), \dots, a_k(G))$ is called the *probability profile* of G .

Definition 12. Given a partition \mathcal{P} and a probability profile $\mathbf{a} \in [0, 1]^k$, define the \mathcal{P} -entropy of \mathbf{a} as

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

Next, using standard Stirling’s approximations we relate the \mathcal{P} -entropy of a probability profile with the entropy of an edge-profile \mathbf{m} .

Lemma 13. 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:

$$\text{ENT}(\mathbf{m}) = H_{\mathcal{P}}(\mathbf{a}) + \gamma(n)$$

where $0 \leq \gamma(n) \leq k \log n$.

The quality of approximation increases as p_i, m_i become larger. Additionally, assuming that $p_i(n) \rightarrow \infty$ as n grows, we get that any point in $[0, 1]^k$ can be well approximated by a probability profile \mathbf{a} , corresponding to some $\mathbf{m} \in \mathcal{M}_{\mathcal{P}}$. Furthermore, continuity properties of the entropy function imply that for understanding the geometry of max-entropy points $M_{\mathcal{P}}^{\epsilon}(S)$, we can instead work with the “smoothed” version using the probability profiles.

Definition 14. Given a \mathcal{P} -symmetric set S , let $\mathbf{a}(S)$ be the set of valid probability-profiles, define the MAX-ENT optimization problem indexed by (\mathcal{P}, S) as:

$$\begin{aligned} \max_{\mathbf{b} \in [0,1]^k} \quad & H_{\mathcal{P}}(\mathbf{b}) \\ \text{s.t.} \quad & \mathbf{b} \in \mathbf{a}(S) \end{aligned}$$

Definition 15. A \mathcal{P} -symmetric set S is convex iff $\mathbf{a}(S)$ is convex, in the sense that for every two points $\mathbf{b}, \mathbf{c} \in \mathbf{a}(S)$, every point in the set $\{w = t\mathbf{b} + (1-t)\mathbf{c}, \forall t \in [0, 1]\} \cap \mathcal{A}_S$ (counterpart of \mathcal{M}_S) also belongs in $\mathbf{a}(S)$.

At this point we are ready to introduce the concept of a *Convex Random Graph*

Definition 16. Given a partition \mathcal{P} and a \mathcal{P} -symmetric convex set S , the random variable $G \sim \text{CRG}_{\mathcal{P}}(S)$ is called a *Convex random graph*.

By restricting our attention to convex sets S , we have the following important consequences:

1. **Uniqueness:** the concavity of \mathcal{P} -entropy and the convexity of $\mathbf{a}(S)$ imply [16] that the MAXENT problem has a unique solution.
2. **Closedness:** since the intersection of convex sets is a convex set, convex random graphs are closed with respect to intersections. Specifically, if we have specified two \mathcal{P} -symmetric convex sets S_1 and S_2 (two different desired properties) their intersection $S_1 \cap S_2$ is also \mathcal{P} -symmetric and convex. This observation is important because it opens the prospect of *engineering*.

The first point permits us to associate with each non-empty convex set S , a unique probability profile $\mathbf{q}^* = \mathbf{q}^*(S)$ coming from the solution of the MAXENT optimization problem for $\mathbf{a}(S)$.

Definition 17. Given a partition \mathcal{P} and a convex set S , the random graph formed by including each edge in P_i with probability $q_i^*(S)$ for all $i \in [k]$, is called the $(\mathcal{P}, \mathbf{q}^*)$ -Edge Block Model and is symbolically expressed as $G \sim \text{EBM}_{\mathcal{P}}(\mathbf{q}^*)$. The corresponding probability distribution will be denoted as $\mathbb{P}_{\mathbf{q}^*}(\cdot)$.

We will show in Section 8 that properties of Convex Random Graphs are closely associated to those of Edge Block Models, and hence that our satisfies the requirement of tractability that we initially set.

7 Concentration of the Edge-Profile

In this section we prove Theorem 1. We aim to characterize the vectors $\mathbf{m} \in \mathbf{m}(S)$ that will be observed with high probability when sampling from the uniform measure on S . In order to achieve that, we need to study the induced probability distribution on $\mathbf{m}(S)$ by the uniform measure on S . This distribution, given in (11), is expressed in terms of the entropy $\text{ENT}(\mathbf{m})$ of \mathbf{m} , and the partition function $Z(S) = |S|$:

$$\mathbb{P}_S(\mathbf{m}) = \frac{1}{|S|} e^{\text{ENT}(\mathbf{m})}$$

The general idea is to identify a high-probability set \mathcal{L} by integrating this probability distribution around the entropy-maximizing profile \mathbf{m}^* . There are two obstacles to overcome in order to carry out this approach successfully: i) the partition function $|S|$ is unknown, meaning that the exact probability distribution $\mathbb{P}_S(\mathbf{m})$ is also unknown, ii) performing a complicated integration over the discrete space $\mathbf{m}(S)$ in a way such that the component-wise maximum distance from \mathbf{m}^* in the set \mathcal{L} is “small”. Requiring small component-wise distance is motivated by the fact that accurate knowledge of \mathbf{m} for each component can be used through Proposition 10, to control the overall distribution \mathbb{P}_S . Our strategy to resolve the above obstacles is:

1. **Polarization:** approximate(lower bound) the log-partition function $\log |S|$ by $\text{ENT}(\mathbf{m}^*)$, the contribution coming only from the entropic optimal edge-profile \mathbf{m}^* :

$$\log \mathbb{P}_S(\mathbf{v}) = \text{ENT}(\mathbf{m}) - \log(|S|) \leq \text{ENT}(\mathbf{m}) - \text{ENT}(\mathbf{m}^*) \quad (15)$$

2. **Smoothing:** approximate $\text{ENT}(\mathbf{m})$ by the smooth counterpart $H_{\mathcal{P}}(\mathbf{a})$ using Stirling’s approximation.
3. **Distance bounds:** exploiting concavity and differentiability of entropy, lower bound the rate that entropy decays as a function of the component-wise distance from the maximizer \mathbf{a}^* (respectively \mathbf{m}^*). This is done by using second-order Taylor approximations.
4. **Union bound:** integrate the obtained bounds outside the set of interest by showing that even if all “bad” vectors were placed right at the boundary of the set, where the lower bound is smallest, the total probability mass would be exponentially small.

We start by approximating entropy via the \mathcal{P} -entropy.

Lemma 18. *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:*

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

where $0 \leq \gamma(n) \leq k \log n$ is a term that approaches zero 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:

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

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

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

where $0 \leq \gamma(n) \leq k \log n$ □

Next, using the Taylor remainder theorem:

Theorem 19 (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:*

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{a})^T (\mathbf{b} - \mathbf{a}) + \frac{1}{2} (\mathbf{b} - \mathbf{a})^T \nabla^2 f(\mathbf{z}) (\mathbf{b} - \mathbf{a}) . \quad (16)$$

and the connection with \mathcal{P} -entropy, we obtain geometric estimates on the decay of entropy around \mathbf{m}^* :

Lemma 20 (L_2 distance bounds). *If \mathbf{m}^* is the unique maximizer and $\mathbf{w} \in \mathbf{m}(S)$, then*

$$\text{ENT}(\mathbf{w}) - \text{ENT}(\mathbf{m}^*) \leq - \sum_{i=1}^k \frac{(w_i - m_i^*)^2}{\max\{\tilde{m}_i^*, \tilde{w}_i\}} + 3k \log n \quad (17)$$

where $\mathbf{m}_i^* = \min\{m_i^*, p_i - m_i^*\}$ (respectively \tilde{w}_i), denotes the ‘‘emptiness’’ of a block.

Proof. Invoking Lemma 18, we rewrite the difference in entropy as a difference in \mathcal{P} -entropy, where \mathbf{a}^* is the probability profile of the maximizer and \mathbf{b} of \mathbf{w} :

$$\text{ENT}(\mathbf{w}) - \text{ENT}(\mathbf{m}^*) \leq H_{\mathcal{P}}(\mathbf{b}) - H_{\mathcal{P}}(\mathbf{a}^*) + 3k \log n$$

Here, we have additionally dealt with the subtle but trivial issue that the probability profile corresponding to \mathbf{m}^* can be slightly different (at most $1/p_i$ in each component) than \mathbf{a}^* . This issue arises out of the fact that $p_i a_i^*$ is not necessarily in $\{0, \dots, p_i\}$ and thus does not correspond to a valid edge profile per-se. It can easily be addressed by using a first-order Taylor approximation and adds a $2k \log n$ term to the bound.

Convexity of the domain and differentiability of the \mathcal{P} -entropy provide the necessary conditions to use the Taylor Remainder Theorem. We proceed with writing the expressions for partial derivatives of $H_{\mathcal{P}}$.

$$\partial_i H_{\mathcal{P}}(\mathbf{a}^*) = -p_i \log \left(\frac{a_i^*}{1 - a_i^*} \right) \quad (18)$$

$$\partial_{ii}^2 H_{\mathcal{P}}(\mathbf{z}) = -p_i \left(\frac{1}{1 - z_i} + \frac{1}{z_i} \right) , \quad (19)$$

while $\partial_{ij}^2 f = 0$ for $i \neq j$, due to separability of the function $H_{\mathcal{P}}$ and \mathbf{z} is a point between \mathbf{a}^* and \mathbf{b} . The Taylor Remainder formula, now reads:

$$H_{\mathcal{P}}(\mathbf{b}) - H_{\mathcal{P}}(\mathbf{a}^*) = \nabla H_{\mathcal{P}}(\mathbf{a}^*) \cdot (\mathbf{b} - \mathbf{a}^*) - \sum_{i=1}^K p_i (b_i - a_i^*)^2 \left(\frac{1}{1 - z_i} + \frac{1}{z_i} \right) \quad (20)$$

Since, \mathbf{a}^* is the unique solution to the sc 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 $\tilde{z}_i = \min\{z_i, 1 - z_i\}$ (expressing the fact that binary entropy is symmetric around $1/2$) and use the trivial bound $\tilde{z}_i \leq \max\{\tilde{a}_i^*, \tilde{b}_i\}$. Thus,

$$H_{\mathcal{P}}(\mathbf{b}) - H_{\mathcal{P}}(\mathbf{a}^*) \leq - \sum_{i=1}^K p_i (b_i - a_i^*)^2 \frac{1}{\tilde{z}_i} \leq - \sum_{i=1}^K p_i \frac{(b_i - a_i^*)^2}{\max\{\tilde{a}_i^*, \tilde{b}_i\}} . \quad (21)$$

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

$$H_{\mathcal{P}}(\mathbf{b}) - H_{\mathcal{P}}(\mathbf{a}^*) \leq - \sum_{i=1}^K \frac{(w_i - m_i^*)^2}{\max\{\tilde{m}_i^*, \tilde{w}_i\}}. \quad (22)$$

and concludes the proof. We note that for most cases we have that $\tilde{z}_i = z_i$, i.e. a block is half-empty. \square

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

Proposition 21. *The number of distinct edge-profiles $|\mathbf{m}(S)|$ is bounded by $|\mathcal{M}_{\mathcal{P}}| \leq e^{2k \log n}$.*

Proof. Assuming that no constraint is placed upon \mathbf{m} by S , then $\mathbf{m}(S) = \mathcal{M}_{\mathcal{P}}$. This number is equal to the product of $[p_i + 1] \leq n^2$ as there are at most $\binom{n}{2}$ edges within a block. Multiplying the last bound we get the statement. \square

Definition 22. *Given a partition \mathcal{P} and a \mathcal{P} -symmetric set S , define:*

$$\mu(S) = \min_{i \in [k]} \{m_i^* \wedge (p_i - m_i^*)\} \quad (23)$$

$$\lambda(S) = \frac{5k \log n}{\mu(S)} \quad (24)$$

$$r(S) = \frac{\lambda + \sqrt{\lambda^2 + 4\lambda}}{2} \quad (25)$$

the thickness, condition number and resolution of the convex set S .

The exact motivation and definition of those quantities will be apparent in the proof of Theorem 1. Here we only give a high level motivation. The *thickness* captures the contribution to the total entropy of the least "occupied" block of edges. On the other hand, the *condition number* intuitively expresses how "singular" the set S is, i.e. whether or not there is a sharp region. Lastly, the *resolution* provides the minimum scale that deviations from the optimum are discernible and corresponds to flatness or sharpness of the geometry around the optimum. We are now ready to prove Theorem 1 following the strategy outlined in the beginning of the section.

Theorem 23. *Given a \mathcal{P} -symmetric set S with $\lambda(S) < 1$. Let $G \sim \text{CRG}_{\mathcal{P}}(S)$ and $\epsilon > r(S)$, then:*

$$\mathbb{P}_S (|\mathbf{m}(G) - \mathbf{m}^*| \leq \epsilon \tilde{\mathbf{m}}^*) \geq 1 - \exp \left[-\mu(S) \left(\frac{\epsilon^2}{1 + \epsilon} - \lambda(S) \right) \right] \quad (26)$$

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

Proof of Theorem 1. 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 $\mathcal{L}_{\epsilon}(\mathbf{m}^*) \triangleq \{G \in S : |\mathbf{m}(G) - \mathbf{m}^*| \leq \epsilon \mathbf{m}^*\}$. We are going to show that $\mathbb{P}_S(\mathcal{L}_{\epsilon}^c(\mathbf{m}^*)) \rightarrow 0$ "exponentially" fast and thus provide localization of the edge profile within a scale ϵ for each coordinate. To that end, we write:

$$\mathbb{P}_S(\mathcal{L}_{\epsilon}^c(\mathbf{m}^*)) = \sum_{\mathbf{w} \in \mathcal{L}_{\epsilon}^c(\mathbf{m}^*)} \mathbb{P}_S(\mathbf{w}) \leq \sum_{\mathbf{w} \in \mathcal{L}_{\epsilon}^c(\mathbf{m}^*)} \exp [\text{ENT}(\mathbf{w}) - \text{ENT}(\mathbf{m}^*)] \quad (27)$$

where we have used (15) and added the contribution of points outside of $\mathbf{m}(S)$. At this point, we are going to leverage the lower bound for the decay of entropy away from \mathbf{m}^* . This is done by first performing a union bound, i.e. considering that all points in $\mathcal{L}_\epsilon^c(\mathbf{m}^*)$ are placed on the least favorable such point \mathbf{w}^* . Since, we are requiring coordinate-wise concentration, such point would differ from the optimal vector only in one-coordinate, and in particular should be the one that minimizes our lower bound. Any such vector $\mathbf{w} \in \mathcal{L}_\epsilon^c(\mathbf{m}^*)$, would have at least one coordinate $i \in [k]$ such that $|w_i - m_i^*| = \epsilon m_i^*$. By Lemma 17, we get

$$\text{ENT}(\mathbf{w}) - \text{ENT}(\mathbf{m}^*) \leq -\frac{\epsilon^2(m_i^*)^2}{(1+\epsilon)m_i^*} + 3k \log n = -\frac{\epsilon^2}{(1+\epsilon)}m_i^* + 3k \log n \quad (28)$$

using the facts that $\max\{\tilde{m}_i, \tilde{w}_i\} \leq \tilde{m}_i + \tilde{w}_i \leq (1+\epsilon)m_i^*$. Now, by definition $\mu(G) \leq \tilde{m}_i^*$ for all $i \in [k]$, and so a vector \mathbf{w}^* that minimizes the bound is such that $\text{ENT}(\mathbf{w}^*) - \text{ENT}(\mathbf{m}^*) \leq -\frac{\epsilon^2}{(1+\epsilon)}\mu(G) + 3k \log n$. We perform the union bound by using $|\mathcal{L}_\epsilon^c(\mathbf{m}^*)| \leq |\mathcal{M}_{\mathcal{P}}| \leq \exp(2k \log n)$ from Proposition 21:

$$\mathbb{P}_S(\mathcal{L}_\epsilon^c(\mathbf{m}^*)) \leq |\mathcal{L}_\epsilon^c(\mathbf{m}^*)| \cdot \mathbb{P}_S(\mathbf{w}^*) \quad (29)$$

$$\leq \exp\left[-\frac{\epsilon^2}{(1+\epsilon)}\mu(G) + 5k \log n\right] \quad (30)$$

$$\leq \exp\left[-\mu(G) \left(\frac{\epsilon^2}{1+\epsilon} - \frac{5k \log n}{\mu(G)}\right)\right] \quad (31)$$

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, whereas $\lambda(S) < 1$ is a requirement that makes potentially possible that also $\epsilon < 1$. \square

Comments on the Proof The theorem shows that there is a bound of $5k \log n$ on the “minimum” number of edges $\mu(S)$ that any block should have. Through the proof, one can identify three sources of error: Stirling’s approximation, integration over $\mathbf{m}(S)$ and approximation of the partition function by the entropy of the most entropic point. In fact, the first two sources of error can be reduced significantly. In particular, one might use higher order Stirling expansions and higher order Taylor approximation for the entropy. Additionally, instead of taking the simple union bound one could define a sequence of “shells” around the optimum, in order to gain by being able to estimate better the number of classes between shells and also by improving the lower bound on the decrease of entropy each time. Unfortunately, although, both strategies work they only lead to an improvement in the leading constant, while unnecessarily complicating the presentation of the proof. The reason is that the main source of error comes from the approximation of the log-partition function by the entropy of the optimal profile. In order to improve this approximation we would need to integrate near the optimum. However, in complete generality, convexity only guarantees that there is at least one direction over which we can integrate. But even then, performing this integration would not yield an improvement as there are only polynomially many classes (n^2) along any single direction.

8 Convex Random Graphs and Edge Block Models

In this section, we leverage the concentration theorem to relate properties of Convex Random Graphs to those of Edge Block Models.

8.1 Probabilistic Clumping of Random Graphs

We begin by constructing a well-known coupling between $G(n, p)$ and $G(n, m)$.

Generating $G(n, m)$ and $G(n, p)$. Recalling the boolean representation of a graph as a string \mathbf{x} in the hypercube $H_n = \{0, 1\}^{\binom{n}{2}}$, we can generate a $G(n, m)$ random graph by: i) assigning to each (edge) bit x_i a uniform random variable $U_i \sim U[0, 1]$, ii) sorting the sequence $U_1, \dots, U_{\binom{n}{2}}$ and selecting the m edges that correspond to the m smallest random variables. That is, given $U_i, i \in [\binom{n}{2}]$, the edge set is a deterministic function $f_{n,m}(U)$ of U . Similarly, one can generate a $G(n, p)$ random graph by i) assigning again uniform random variables U_i for each bit x_i ii) selecting all edges such that $U_i \leq p$. Once more, the edge set is a deterministic function $g_{n,p}(U)$ of U .

We are interested in relating three random graphs $G^- \sim G(n, p_-)$, $G \sim G(n, m)$ and $G^+ \sim G(n, p_+)$. This can be achieved by leveraging the previous facts and defining a common probability space where events(inclusion-exclusion of edges) take place.

Coupling Consider the the probability distribution $\mathbb{P}_U(\cdot)$ on $[0, 1]^{\binom{n}{2}}$ of $\binom{n}{2}$ i.i.d $U[0, 1]$ random variables. Each element of the sample space consists of a vector $\mathbf{u} = (u_1, \dots, u_{\binom{n}{2}}) \in [0, 1]^{\binom{n}{2}}$. In this space the edge sets of random graphs become deterministic functions of U . Furthermore, for each model we both have that the marginal probability distribution are exactly the same as the original distributions, but now we can relate events happening for one random graph to another.

In particular, the same coupling can be constructed for every setting where we sample a subset S of m items out of a set A (sampling without replacement) denoted as $S \sim WR(A, m)$, and when we include each item in the set A with probability p (sampling with replacement) denoted as $S \sim R(A, p)$.

Lemma 24. *Given m and a set A with $|A| = n$, fix an $\delta > 0$ and define $p_{\pm} = \frac{m}{(1 \mp \delta)n}$. Then for $S \sim WR(A, m)$, $S^{\pm} \sim R(A, p_{\pm})$, we have that $S_- \subset S \subset S_+$ with probability at least $1 - 2 \exp\left(-\frac{\delta^2}{3(1+\delta)}m\right)$.*

Proof. By the construction of sets S, S_{\pm} in order to prove the lemma it suffices to show that: $|S_-| \leq |S| \leq |S_+|$ with large enough probability. To that end, define the bad events

$$B_- = \{u \in [0, 1]^{|A|} : \sum_{i=1}^{|A|} \mathbb{I}(u_i \leq p_-) > m\}$$

$$B_+ = \{u \in [0, 1]^{|A|} : \sum_{i=1}^{|A|} \mathbb{I}(u_i \leq p_+) < m\}$$

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 then) the expectation np_{\pm} .By employing standard Chernoff bounds, we get:

$$\mathbb{P}_U(B_-) = \mathbb{P}_U(X_- > m) = \mathbb{P}_U(X > (1 + \delta)np_-) \leq \exp\left(-\frac{\delta^2}{3(1 + \delta)}m\right)$$

$$\mathbb{P}_U(B_+) = \mathbb{P}_U(X_+ < m) = \mathbb{P}_U(X < (1 - \delta)np_+) \leq \exp\left(-\frac{\delta^2}{2(1 - \delta)}m\right)$$

The proof is concluded through the use of union bound:

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

□

Using, this simple lemma, we can prove the *sandwich theorem*.

Theorem 25. Let $G^\pm \sim \text{EBM}_{\mathcal{P}}(\mathbf{q}^\pm)$ and $G \sim \text{CRG}_{\mathcal{P}}(S)$, where \mathbf{q}^\pm is defined by $q_i^\pm = \frac{1-\epsilon}{1\mp\epsilon} q_i^*$ for all $i \in [k]$, and $\mathbf{q}^*(S)$ is as before. For sets $S \subset \mathcal{G}_n$, satisfying the conditions of Theorem :

$$G^- \subset G \subset G^+ \quad (32)$$

with probability at least $1 - 2 \exp \left[-\mu(S) \left(\frac{\epsilon^2}{12} - \lambda(S) \right) \right]$ and $\epsilon < 1/2$.

Proof. To prove our theorem, we first observe that if we condition on the edge profile \mathbf{m} of the random graph $G \sim \text{CRG}_{\mathcal{P}}(S)$, by Proposition 10 the probability distribution factorizes in $G(n, m)$ -like (sampling without replacement) distributions for each block in the partition \mathcal{P} . For each block $i \in [k]$ we can construct a coupling between $E_i \sim WR(P_i, m_i)$ and $E_i^\pm \sim R(P_i, p_i^\pm)$ by using again i.i.d. uniform random variables. Then, we can apply Lemma 24, to obtain bounds on the probability that the relationship $E_i^- \subset E_i \subset E_i^+$ does not hold for a given block. Using the union bound, we then obtain an estimate of the probability (conditional on the edge profile) that the property holds across blocks. The final step then, is to use the concentration theorem to prove, that for all edge-profiles in the concentration region, this property holds with high probability.

Concretely, define B_i the event that the i -th block does not satisfy the property $E_i^- \subset E_i \subset E_i^+$. We rewrite the event of interest by considering the complement, i.e. $\mathbb{P}_U(G^- \subset G \subset G^+) = 1 - \mathbb{P}_U(\cup B_i)$. Conditioning on the edge profile gives:

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

where 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. The second inequality, is derived by upper bounding the probability of the bad event by the most favorable such edge-profile. Lastly, the last inequality is due to an application of the union bound. Applying Theorem 1 we get a bound on the first term and then invoking Lemma 24 we get a bound for each of the term in the sum:

$$\mathbb{P}_U(\cup B_i) \leq \exp \left[-\mu(S) \left(\frac{\epsilon^2}{1+\epsilon} - \lambda(S) \right) \right] + 2 \max_{\mathbf{v} \in L_\epsilon(\mathbf{m}^*)} \left[\sum_{i=1}^k \exp \left(-\frac{\epsilon^2}{3(1+\epsilon)} v_i \right) \right]$$

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

$$\mathbb{P}_U(\cup B_i) \leq \exp \left[-\mu(S) \left(\frac{\epsilon^2}{1+\epsilon} - \lambda(S) \right) \right] + 2k \exp \left[-\frac{\epsilon^2(1-\epsilon)}{3(1+\epsilon)} \mu(S) \right] \quad (33)$$

$$\leq \exp \left[-\mu(S) \left(\frac{\epsilon^2}{1+\epsilon} - \lambda(S) \right) \right] + \exp \left[-\mu(S) \left(\frac{\epsilon^2(1-\epsilon)}{3(1+\epsilon)} - \frac{\log(2k)}{\mu(S)} \right) \right] \quad (34)$$

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

8.2 Approximate Independence for Local Events

The concentration and sandwich theorems provide a global view of Convex random graphs. Here, we aim to refine this view and provide accurate estimates of the probabilities under \mathbb{P}_S of local events, i.e. concerning a few edges. This is motivated by the fact that in many occasions people are interested in what happens in small regions of the graph, like the appearance of a certain subgraph(e.g. triangles, cliques, cycles) or the expected degrees of nodes. Towards that end, we are going to exploit once more the duality between $G(n, m)$ and $G(n, p)$.

We start off by defining what is a local event. Consider any disjoint sets of edges $I_i, O_i \subset P_i$ and define the event $A_i = \{G \in \mathcal{G}_n : I_i \subset E(G) \text{ and } O_i \cap E(G) = \emptyset\}$ for $i \in [k]$. Our definition of local events will involve the number of edges appearing in the events, as well as the probability of the events. The first consideration is naturally motivated by the $G(n, m)$ -like decomposition of the measure and the control we have over its parameters(edge-profile). The second requirement is due to the fact that we can estimate the probabilities only through the concentration theorem and, thus, events that have much smaller probabilities than the error in the concentration theorem are masked by it. We proceed with the definition of *local events*.

Definition 26. Assume a \mathcal{P} -symmetric convex set S , and let $\mathbf{m}^*(S), \mathbf{q}^*(S)$ denote the optimal edge(probability) profile. Fix $\epsilon > r(S)$ and $\gamma > 0$, an event A_i is (ϵ, γ) -local if:

$$|I_i \cup O_i| \leq \epsilon \cdot m_i^* \quad (35)$$

$$\frac{q_i^*}{1 - q_i^*} \leq \gamma \quad (36)$$

where $r(S)$ denotes as before the resolution of S . Furthermore, a composite event $A = \cap_{i=1}^k A_i$ is (ϵ, γ) -local if additionally:

$$\mathbb{P}_{\mathbf{q}^*}(A) \geq \exp \left[-\mu(S) \left(\frac{\epsilon^2}{1 + \epsilon} - \lambda(S) \right) \right] \cdot \epsilon \quad (37)$$

The requirement (36) that the ratio is bounded implies that $q_i^* \leq 1/(1 + \gamma)$ and essentially is needed in order to consider events A_i where some edges might not be present(otherwise the probability would be arbitrary close to 0 even when O_i consisted of a single edge). Let $|I_i| = s_i$ and $|O_i| = n_i - s_i$ be the cardinalities of the sets appearing in A_i . We start with a lemma that provides estimates about the conditional probability distribution within each block:

Lemma 27. For a \mathcal{P} -symmetric convex set S with $\lambda(S) < 1$, let $\epsilon > r(S)$ and consider $\mathbf{v} \in \mathcal{L}_\epsilon(\mathbf{m}^*)$ (in the concentration region). For a (ϵ, γ) -local event A_i , we have:

$$(1 - 2\epsilon)^{s_i} [1 - 2\epsilon\gamma]^{n_i - s_i} \leq \frac{\mathbb{P}_S(A_i|\mathbf{v})}{P_{\mathbf{q}^*}(A_i)} \leq (1 - \epsilon)^{-n_i} (1 + \gamma\epsilon)^{n_i - s_i} \quad (38)$$

where $(q_i^*)^{s_i} (1 - q_i^*)^{n_i - s_i} = P_{\mathbf{q}^*}(A_i)$ is the probability of the event under the $EBM_{\mathcal{P}}(\mathbf{q}^*)$.

Proof. We start off by righting an exact expression for $\mathbb{P}_S(A_i|\mathbf{v})$:

$$\mathbb{P}_S(A_i|\mathbf{v}) = \frac{\binom{p_i - n_i}{v_i - s_i}}{\binom{p_i}{v_i}} = \frac{(v_i)_{s_i} (p_i - v_i)_{n_i - s_i}}{(p_i)_{n_i}} \quad (39)$$

where $(x)_n$ denotes the descending factorial, for which we get the following easy bounds $(x - n)^n \leq x(x - 1) \dots (x - n + 1) \leq x^n$. First, we obtain a lower bound on $\mathbb{P}_S(A_i|\mathbf{v})$:

$$\mathbb{P}_S(A_i|\mathbf{v}) \geq \frac{(v_i - s_i)^{s_i} (p_i - v_i - n_i + s_i)^{n_i - s_i}}{p_i^{n_i}} \quad (40)$$

$$\geq \left[(1 - 2\epsilon) \frac{m_i^*}{p_i} \right]^{s_i} \left[1 - (1 + 2\epsilon) \frac{m_i^*}{p_i} \right]^{n_i - s_i} \quad (41)$$

$$\geq (1 - 2\epsilon)^{s_i} (q_i^*)^{s_i} (1 - q_i^*)^{n_i - s_i} \left[1 - \frac{q_i^*}{1 - q_i^*} 2\epsilon \right]^{n_i - s_i} \quad (42)$$

$$\geq \mathbb{P}_{\mathbf{q}^*}(A_i) (1 - 2\epsilon)^{s_i} [1 - 2\epsilon\gamma]^{n_i - s_i} \quad (43)$$

The second inequality was derived using $\mathbf{v} \in \mathcal{L}_\epsilon(\mathbf{m}^*)$ and $n_i \leq \epsilon m_i^*$, while the third by recalling that $q_i^* = m_i^*/p_i$. To get the last inequality, we use the assumption $q_i^*/(1 - q_i^*) < \gamma$. Proceeding in the same manner we get the upper bound:

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

□

The idea now is to combine those partial conditional estimates for all classes and use the concentration theorem to get rid of conditioning.

Theorem 28. *Given a \mathcal{P} -symmetric convex set $S \subset \mathcal{G}_n$ with $\mu(S) \gg (5k \log n)^2$, define $\hat{\epsilon} \triangleq \sqrt{c^* \lambda(S)}$ where $c^* \approx 2$ is explicitly defined. For an $(\hat{\epsilon}, \gamma)$ -local event A involving at most $m \leq \frac{\sqrt[4]{\mu(S)}}{\sqrt{c^* 5k \log n}}$ edges, it holds:*

$$\left| \frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{q}^*}(A)} - 1 \right| \leq \frac{8\gamma}{\sqrt[4]{\mu(S)}} \quad (44)$$

where $\mathbb{P}_{\mathbf{q}^*}(A)$ denotes the probability under $\text{EBM}_{\mathcal{P}}(\mathbf{q}^*)$.

Proof. To prove the lemma we use again the usual strategy of conditioning. In particular, we obtain initial upper and lower bounds by utilizing Theorem 1 and conditioning. Let $Q_S(\epsilon)$ be the ‘‘error’’ of the concentration theorem:

$$(1 - Q_S(\epsilon)) \min_{\mathbf{v} \in \mathcal{L}_\epsilon(\mathbf{m}^*)} \mathbb{P}_S(A|\mathbf{v}) \leq \mathbb{P}_S(A) \leq \mathbb{P}_S(\mathcal{L}_\epsilon^c(\mathbf{m}^*)) + \max_{\mathbf{v} \in \mathcal{L}_\epsilon(\mathbf{m}^*)} \mathbb{P}_S(A|\mathbf{v})$$

By dividing with $\mathbb{P}_{\mathbf{q}^*}(A)$, we get:

$$(1 - Q_S(\epsilon)) \min_{\mathbf{v} \in \mathcal{L}_\epsilon(\mathbf{m}^*)} \frac{\mathbb{P}_S(A|\mathbf{v})}{\mathbb{P}_{\mathbf{q}^*}(A)} \leq \frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{q}^*}(A)} \leq \frac{\mathbb{P}_S(\mathcal{L}_\epsilon^c(\mathbf{m}^*))}{\mathbb{P}_{\mathbf{q}^*}(A)} + \max_{\mathbf{v} \in \mathcal{L}_\epsilon(\mathbf{m}^*)} \frac{\mathbb{P}_S(A|\mathbf{v})}{\mathbb{P}_{\mathbf{q}^*}(A)}$$

Due to Proposition 10 and Lemma 27, we have the following bounds for the ratios involving conditioning on the edge-profile:

$$(1 - 2\epsilon)^m [1 - 2\epsilon\gamma]^m \leq \prod_{i=1}^k \frac{\mathbb{P}_S(A_i|\mathbf{v})}{\mathbb{P}_{\mathbf{q}^*}(A_i)} \leq (1 - \epsilon)^{-m} (1 + \gamma\epsilon)^m \quad (45)$$

where $m = \sum n_i$ is the total number of edges. Moreover, by definition of the ϵ -local event, we get that $\frac{\mathbb{P}_S(\tilde{\mathcal{L}}_\epsilon(\mathbf{m}^*))}{\mathbb{P}_{\mathbf{q}^*}(A)} \leq \epsilon$. Now, setting $\hat{\epsilon} = \sqrt{c^* \lambda(S)}$, where $c^* = 2(1 + \lambda + \sqrt{\lambda^2 + 2\lambda}) \approx 2$, such that $\hat{\epsilon}^2/(1 + \hat{\epsilon}) - \lambda = \lambda$ and recalling that $\mu(S)\lambda = 5k \log n$, gives:

$$\frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{q}^*}(A)} \geq \left(1 - \frac{1}{n^{5k}}\right) (1 - \hat{\epsilon})^m [1 - 2\gamma\hat{\epsilon}]^m \quad (46)$$

$$\frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{q}^*}(A)} \leq \hat{\epsilon} + (1 - \hat{\epsilon})^{-m} (1 + \gamma\hat{\epsilon})^m \quad (47)$$

For the lower bound we use the inequality $(1 - x)^y \geq 1 - xy$, $\forall x < 1$, and obtain:

$$\frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{q}^*}(A)} \geq \left(1 - \frac{1}{n^{5k}}\right) (1 - \hat{\epsilon}m) [1 - 2\gamma\hat{\epsilon}m] \quad (48)$$

$$\geq 1 - 8\gamma\hat{\epsilon}m \quad (49)$$

The last line follows by expanding the product and bounding from below each term in the sum by the smallest such term. To get the upper bound we proceed similarly. We use the inequality $(1 - x)^y \geq 1 - xy$, $\forall x < 1$ to simplify the fraction in the denominator. Further, for $\hat{\epsilon}m \rightarrow 0$ we have that $(1 + \gamma\hat{\epsilon})^m \leq 1 + 2\gamma\hat{\epsilon}m$ and $\frac{1}{1 - \hat{\epsilon}m} \leq 1 + 2\hat{\epsilon}m$ for adequately large n . Consequently:

$$\frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{q}^*}(A)} \leq \hat{\epsilon} + (1 + 2m\hat{\epsilon})(1 + 2\gamma m\hat{\epsilon}) \quad (50)$$

$$\leq 1 + 6\gamma\hat{\epsilon} \quad (51)$$

Putting everything together, we obtain that:

$$\left| \frac{\mathbb{P}_S(A)}{\mathbb{P}_{\mathbf{q}^*}(A)} - 1 \right| \leq 8\gamma\hat{\epsilon}m$$

Finally, by assumption $\hat{\epsilon}m \leq \frac{1}{\sqrt[4]{\mu(S)}}$ and, hence, the theorem is concluded. \square

Definition 29. A function $f : \mathcal{G}_n \rightarrow \mathbb{R}$ is called (ϵ, γ) -local for a \mathcal{P} -symmetric set S if $\forall x \in \mathbb{R}$ the event $B_x \triangleq \{G \in \mathcal{G}_n : f(G) \leq x\}$ is (ϵ, γ) -local.

Corollary 30. Under the conditions of Theorem 3, for any (ϵ, γ) -local function $f : \mathcal{G}_n \rightarrow \mathbb{R}$, it holds that:

$$|\mathbb{E}_S[f(G)] - \mathbb{E}_{\mathbf{q}^*}[f(G)]| \leq \frac{8\gamma}{\sqrt[4]{\mu(S)}} \mathbb{E}_{\mathbf{q}^*}[|f(G)|] \quad (52)$$

Proof. Observe that since the function is defined on the discrete space \mathcal{G}_n it takes at most $2^{\binom{n}{2}}$ values. Let $I^+ = \{x \in \mathbb{R} : \exists G \in \mathcal{G}_n, f(G) = x > 0\}$ and I^- be (respectively) the positive and negative values that f can take. We simply multiply (44) by $\mathbb{P}_{\mathbf{q}^*}(A)$, and integrate it over the two sets. \square

Remark 31. The corollary shows that for function of well-conditioned parts of the graphs, we have accurate control. Moreover, if we assume the existence of higher moments, we get information about the distribution of f . As an example, one can obtain a feel for the distribution of T , the number of triangles in the graphs by calculating higher moments, that would be sums of local functions of more and more edges.

9 Prior Information and Probabilistic Reductions

In the exposition, the introduction of symmetry was presented as a way to endow structure to the otherwise unstructured arbitrary subsets of \mathcal{G}_n . In reality, the introduction of symmetry is a way to incorporate *prior knowledge* about the set of graphs of interest. The question is what kind of information is our notion of symmetry able to encompass?

Equivalence as a Model We promote the notion that a *model* (\mathcal{X}, \asymp) on a space \mathcal{X} is defined through an equivalence relation \asymp on $\mathcal{X} \times \mathcal{X}$. In other words, it captures whether or not two objects contain the same “information” to the eyes of the modeler. In full generality, the equivalence relation induces a partition \mathcal{Q} of \mathcal{X} , where membership to a part is dependent on the full specification of the object. Our symmetry assumptions, define this equivalence relation through:

$$G_1, G_2 \in \mathcal{G}_n : G_1 \asymp G_2 \Leftrightarrow \mathbf{m}(G_1) = \mathbf{m}(G_2)$$

where $\mathbf{m}(G)$ is the edge-profile of the graph. This identification, as we saw, was a consequence of the existence of a partition \mathcal{P} and the assumption of permutation symmetry between edges in the same block. Therefore, in our case the partition \mathcal{Q} of elements of \mathcal{X} , is induced by a (lower order) partition of edges in equivalence classes and, hence, our model can encompass information about *first-order equivalences between edges*.

The model of *Generalized Erdős-Rényi* random graphs is, thus, a *meta-model* where one specifies the prior-information available in terms of the partition of edges. Extensive examples of models, where such information is available, were presented earlier and were unified under the general concept of a *geometric partition* (g, \mathcal{R}) (expressing e.g. membership, similarity, distance). Next we explore the range of properties that are expressible under block-permutation symmetry.

Probabilistic Reductions Given the prior-information specified by a partition \mathcal{P} of edges, the range of sets(properties) expressible are those whose characteristic function $\mathbb{I}_S(G) = \mathbb{I}_{\mathbf{m}(S)}(\mathbf{m}(G))$ is a function only of $\mathbf{m}(G)$ (Proposition 9). Arguably, at first glance this is not a rich family in the case where the partition \mathcal{P} is meaningful (reasonable amount of grouping). However, exploiting the fact that, given the edge profile \mathbf{m} , we have very accurate knowledge of the conditional probability distribution $\mathbb{P}_S(\cdot | \mathbf{m})$ (see Proposition 10, Lemma 24 and Lemma 27) opens the prospect of being able to *probabilistically reduce* the property of interest as a *specification on the set of edge-profiles* $\mathcal{M}_{\mathcal{P}}$.

Definition 32. For a given property(set) S , define its (ϵ, \mathcal{P}) -approximation \hat{S} as:

$$\hat{S}_{\mathcal{P}}(\epsilon) \triangleq \{\mathbf{v} \in \mathcal{M}_{\mathcal{P}} : \mathbb{P}(S | \mathbf{v}) \geq 1 - \epsilon\} \quad (53)$$

and define $\mathbf{m}^{-1}(\hat{S}) = \{G \in \mathcal{G}_n : \mathbf{m}(G) \in \hat{S}\}$ the corresponding set of graphs.

That is, instead of studying S itself, study the pre-image of the set of profiles $\mathbf{v} \in \hat{S}_{\mathcal{P}}(\epsilon)$ such that: if we take a subset of v_i edges at random from each block P_i of the partition, the property holds with probability at least $(1 - \epsilon)$. The hope is that we will be able to obtain a *concise description* of $\hat{S}_{\mathcal{P}}(\epsilon)$ in terms of \mathbf{v} . The latter goal can be achieved through two different avenues:

Phase Transitions: a large part of random graph theory is to establish *threshold functions* for various monotone properties. This theory becomes particular pertinent in our context, since it can be used to derive *tight* and *concise* approximations of graph properties.

- **Tightness:** refers to the fact that the existence of a threshold function, implies that the set $\hat{S}_{\mathcal{P}}(\epsilon)$ is actually a good approximation of the set S . In the sense that, since, conditional on the edge-profile, all graphs with the given profile are equi-probable: i) the number of graphs in the set $\hat{S}_{\mathcal{P}}(\epsilon)$ satisfying the property are an $(1 - \epsilon)$ fraction of $|\hat{S}_{\mathcal{P}}(\epsilon)|$ (positive side of the threshold), ii) The number of graphs in the complement of $\hat{S}_{\mathcal{P}}(\epsilon)$ are a vanishing fraction of all such graphs (negative side of the threshold), iii) for sharp thresholds the number of graphs in between are usually a vanishing fraction compared to both sets. In such cases, the approximation is highly satisfactory.
- **Conciseness:** the existence of threshold function has another important consequence in our context, i.e. the set $\hat{S}_{\mathcal{P}}(\epsilon)$ can be described as $\{g(\mathbf{m}) \geq c_{\epsilon}^*\}$ for some function $g : \mathbb{R}^k \rightarrow \mathbb{R}$. For instance, in the case of the original Erdős-Rényi model, we are given the trivial partition \mathcal{P}_{ER} and aim to study the following sets $S_1 = \{\text{giant component}\}$, $S_2 = \{\text{connectivity}\}$, $S_3 = \{\text{r-qlique}\}$, then:

$$\hat{S}_1 = \left\{ m \in \mathcal{M}_{\mathcal{P}} : m \geq (1 + \epsilon) \frac{n}{2} \right\} \quad (54)$$

$$\hat{S}_2 = \left\{ m \in \mathcal{M}_{\mathcal{P}} : m \geq (1 + \epsilon) \frac{n \log n}{2} \right\} \quad (55)$$

$$\hat{S}_3 = \left\{ m \in \mathcal{M}_{\mathcal{P}} : m \geq \frac{(1 + \epsilon)}{2} n^{2 - \frac{2}{r}} \right\} \quad (56)$$

The above results refer to Erdős-Rényi random graphs where the partition does not play a role, hence, one might think that this is the singular setting where results of the kind we are describing are attainable. We have shown earlier that the same principles apply to far more general models, like the Stochastic Block Model or geometric models. The general point we want to put forward is that *tight* and *concise* approximation to properties of interest are plausible.

References

- [1] Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 793–802. IEEE, 2008.
- [2] Dimitris Achlioptas and Paris Siminelakis. Navigating navigability. *manuscript*, 2014.
- [3] William Aiello, Anthony Bonato, Colin Cooper, J Janssen, and P Prałat. A spatial web graph model with local influence regions. *Internet Mathematics*, 5(1-2):175–196, 2008.
- [4] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2004.
- [5] M. Artin. *Algebra*. Featured Titles for Abstract Algebra Series. Pearson Prentice Hall, 2011.
- [6] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

- [7] Edward A Bender and E Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296–307, 1978.
- [8] Noam Berger, Christian Borgs, Jennifer T Chayes, RM D’souza, and Robert D Kleinberg. Degree distribution of competition-induced preferential attachment graphs. *Combinatorics Probability and Computing*, 14(5/6):697, 2005.
- [9] Shankar Bhamidi, Guy Bresler, and Allan Sly. Mixing time of exponential random graphs. In *Foundations of Computer Science, 2008. FOCS’08. IEEE 49th Annual IEEE Symposium on*, pages 803–812. IEEE, 2008.
- [10] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.
- [11] Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.
- [12] 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.
- [13] C. Borgs, J.T. Chayes, L. Lovsz, V.T. Ss, and K. Vesztergombi. Convergent sequences of dense graphs i: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801 – 1851, 2008.
- [14] Christian Borgs, Jennifer Chayes, Constantinos Daskalakis, and Sebastien Roch. First to market is not everything: An analysis of preferential attachment with fitness. In *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing, STOC ’07*, pages 135–144, New York, NY, USA, 2007. ACM.
- [15] Christian Borgs, Jennifer T Chayes, László Lovász, Vera T Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs ii. multiway cuts and statistical physics. *Ann. of Math.(2)*, 176(1):151–219, 2012.
- [16] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [17] Fan RK Chung and Linyuan Lu. *Complex graphs and networks*, volume 107. American mathematical society Providence, 2006.
- [18] Constantinos Daskalakis, Alexandros G Dimakis, Elchanan Mossel, et al. Connectivity and equilibrium in random games. *The Annals of Applied Probability*, 21(3):987–1016, 2011.
- [19] Maria Deijfen and Willemien Kets. Random intersection graphs with tunable degree distribution and clustering. *Probability in the Engineering and Informational Sciences*, 23(4):661, 2009.
- [20] Sergey N Dorogovtsev and Jose FF Mendes. Evolution of networks. *Advances in physics*, 51(4):1079–1187, 2002.
- [21] Richard Durrett. *Random graph dynamics*, volume 20. Cambridge university press, 2007.
- [22] David Easley and Jon Kleinberg. Networks, crowds, and markets. *Cambridge Univ Press*, 6(1):6–1, 2010.

- [23] Paul Erdős and Alfréd Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.
- [24] Abraham D Flaxman, Alan M Frieze, and Juan Vera. A geometric preferential attachment model of networks. *Internet Mathematics*, 3(2):187–205, 2006.
- [25] Alan Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.
- [26] Alan Frieze, Santosh Vempala, and Juan Vera. Logconcave random graphs. In *Proceedings of the 40th annual ACM symposium on Theory of computing, STOC '08*, pages 779–788, New York, NY, USA, 2008. ACM.
- [27] Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, pages 1141–1144, 1959.
- [28] David F Gleich and Art B Owen. Moment-based estimation of stochastic kronecker graph parameters. *Internet Mathematics*, 8(3):232–256, 2012.
- [29] 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.
- [30] Luca Gugelmann, Konstantinos Panagiotou, and Ueli Peter. Random hyperbolic graphs: Degree sequence and clustering. In *Proceedings of the 39th International Colloquium Conference on Automata, Languages, and Programming - Volume Part II, ICALP'12*, pages 573–585, Berlin, Heidelberg, 2012. Springer-Verlag.
- [31] Martin Haenggi, Jeffrey G Andrews, François Baccelli, Olivier Dousse, and Massimo Franceschetti. Stochastic geometry and random graphs for the analysis and design of wireless networks. *Selected Areas in Communications, IEEE Journal on*, 27(7):1029–1046, 2009.
- [32] Matthew O Jackson. *Social and economic networks*. Princeton University Press, 2010.
- [33] Emmanuel Jacob and Peter Mörters. A spatial preferential attachment model with local clustering. In *Algorithms and Models for the Web Graph*, pages 14–25. Springer, 2013.
- [34] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.
- [35] Gil Kalai and Shmuel Safra. Threshold phenomena and influence: perspectives from mathematics, computer science, and economics. *Computational complexity and statistical physics*, pages 25–60, 2006.
- [36] Jon M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845, August 2000.
- [37] JonM. Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and AndrewS. Tomkins. The web as a graph: Measurements, models, and methods. In Takano Asano, Hideki Imai, D.T. Lee, Shin-ichi Nakano, and Takeshi Tokuyama, editors, *Computing and Combinatorics*, volume 1627 of *Lecture Notes in Computer Science*, pages 1–17. Springer Berlin Heidelberg, 1999.
- [38] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguñá. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, 2010.

- [39] Silvio Lattanzi and D. Sivakumar. Affiliation networks. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 427–434, New York, NY, USA, 2009. ACM.
- [40] 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.
- [41] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2, 2007.
- [42] Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, 2009.
- [43] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.
- [44] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.
- [45] Mohammad Mahdian and Ying Xu. Stochastic kronecker graphs. *Random Structures & Algorithms*, 38(4):453–466, 2011.
- [46] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. *CoRR*, abs/1311.3085, 2013.
- [47] Frank McSherry. Spectral partitioning of random graphs. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 529–537. IEEE, 2001.
- [48] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.
- [49] Alan Mislove, Massimiliano Marcon, Krishna P Gummadi, Peter Druschel, and Bobby Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 29–42. ACM, 2007.
- [50] Mark Newman. *Networks: an introduction*. Oxford University Press, 2010.
- [51] Mark EJ Newman and Michelle Girvan. Mixing patterns and community structure in networks. In *Statistical mechanics of complex networks*, pages 66–87. Springer, 2003.
- [52] C.L.M. Nickel and The Johns Hopkins University. *Random Dot Product Graphs: A Model for Social Networks*. Johns Hopkins University, 2007.
- [53] Mathew Penrose. *Random geometric graphs*, volume 5. Oxford University Press Oxford, 2003.
- [54] Derek de Solla Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5):292–306, 1976.
- [55] Tom Richardson and Rüdiger Leo Urbanke. *Modern coding theory*. Cambridge University Press, 2008.

- [56] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph p^* models for social networks. *Social networks*, 29(2):173–191, 2007.
- [57] Karl Rohe, Sourav Chatterjee, Bin Yu, et al. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.
- [58] Alessandra Sala, Lili Cao, Christo Wilson, Robert Zablit, Haitao Zheng, and Ben Y. Zhao. Measurement-calibrated graph models for social network experiments. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pages 861–870, New York, NY, USA, 2010. ACM.
- [59] Tom AB Snijders. Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40, 2002.
- [60] Gregory Valiant and Tim Roughgarden. Braess’s paradox in large random graphs. *Random Struct. Algorithms*, 37(4):495–515, 2010.
- [61] Alexei Vázquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67(5):056104, 2003.
- [62] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):409–10, 1998.
- [63] Stephen J Young and Edward R Scheinerman. Random dot product graph models for social networks. In *Algorithms and models for the web-graph*, pages 138–149. Springer, 2007.

A Trivial Proofs

Proof of Proposition 9. Fix an $\mathbf{x} \in H_n$ and consider the set $O(\mathbf{x}) \triangleq \{\mathbf{y} \in H_n : \exists \pi \in \Pi_n(\mathcal{P}) \text{ such that } y = \pi(\mathbf{x})\}$ 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) $\pi_x(\mathbf{x}) = (\pi_{x,1}(x_{P_1}), \dots, \pi_{x,k}(x_{P_k})) \in O(\mathbf{x})$, ii) for all $i \in [k]$, $\pi_{x,i}(x_{P_i})$ is a bit-string where all 1’s appear consecutively starting from the first position. Let us 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 $\mathbf{m} = (m_1, \dots, m_k)$. \square

Proof of Proposition 10. Since $G \sim \mathcal{P}(S)$ the distribution of G is by definition uniform on \mathcal{S} . This also means that it is uniform on the subset of graphs having edge profile $\mathbf{m} \in \mathbb{N}^k$ (conditioning). But then:

$$\mathbb{P}_S(A_1 \cap \dots \cap A_k | \mathbf{v}) = \frac{\mathbb{P}_S(A_1 \cap \dots \cap A_k \cap \mathbf{m}(G) = \mathbf{v})}{\mathbb{P}_S(\mathbf{m}(G) = \mathbf{v})} = \frac{|A_1 \cap \dots \cap A_k \cap \mathbf{m}(G) = \mathbf{v}|}{|\mathbf{m}(G) = \mathbf{v}|} \mathbb{I}_{\mathbf{m}(S)}(\mathbf{v})$$

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 $A_i = \{G \in \mathcal{G}_n : I_i \subset E(G) \text{ and } O_i \cap E(G) = \emptyset\}$ 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 \square , we know that exactly v_i edges from P_i are included in G and that we can satisfy the requirements for edges in P_i by selecting any subset of $v_i - |I_i|$ edges out of $P_i \setminus (I_i \cup O_i)$. For convenience set $|P_i| = p_i$, $|I_i| = n_i$, $|O_i \cup I_i| = r_i$, 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 $C_{v_i - n_i}^{p_i - r_i}$. As the constraints imposed are separable, we have:

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

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