

# ASYMPTOTIC QUANTIZATION OF EXPONENTIAL RANDOM GRAPHS

MEI YIN, ALESSANDRO RINALDO, AND SUKHADA FADNAVIS

April 1, 2014

ABSTRACT. We describe the asymptotic properties of the edge-triangle exponential random graph model as the natural parameters diverge along straight lines. We show that as we continuously vary the slopes of these lines, a typical graph drawn from this model exhibits quantized behavior, jumping from one complete multipartite graph to another, and the jumps happen precisely at the normal lines of a polyhedral set with infinitely many facets. As a result, we provide a complete description of all asymptotic extremal behaviors of the model.

## 1. INTRODUCTION

Over the last decades, the availability and widespread diffusion of network data on typically very large scales have created the impetus for the development of new theories for modeling and describing the properties of large random networks. Nonetheless, and despite the vast and rapidly growing body of literature on network analysis (see, e.g., [21, 22, 27, 36, 43] and references therein), the study of the asymptotic behavior of network models has proven rather difficult in most cases. Indeed, because the limiting properties of many such models remain hard to determine, methodologies for carrying out fundamental inferential tasks such as parameter estimation, hypothesis and goodness-of-fit testing with provable asymptotic guarantees have yet to be developed for most network models.

Exponential random graph models [24, 32, 55] form one of the most prominent families of statistical models for random graphs, but also one for which the issue of present lack of understanding of their general asymptotic properties is particularly pressing. These rather general models are exponential families of probability distributions over graphs, whereby the natural sufficient statistics are virtually any functions on the space of graphs that are deemed to capture the features of interest. Such statistics may include, for instance, the number of edges or copies of any finite subgraph, as well as more complex quantities such as the degree sequence or degree distribution, and combinations thereof. As a result, these models are able to capture a wide variety of common network tendencies by representing a complex global structure through a set of tractable local features. They are especially useful when one wants to construct models that resemble observed networks as closely as possible, but without going into details of the specific process underlying network formation. Due to their flexibility and modeling power, exponential random graph models are among the most widespread models in network analysis, with countless applications in multiple disciplines, including social sciences, statistics and statistical mechanics. See, e.g., [44, 50, 51, 53, 56].

Being exponential families with finite support, one might expect exponential random graph models to enjoy a rather simple asymptotic form. Though in fact, virtually all these models are highly nonstandard: as the size of the network increases, they exhibit properties that cannot be described using traditional tools of asymptotic parametric statistics for i.i.d. data, but require instead more sophisticated analysis. Indeed, even for the  $p_1$  model of [32], which

postulates independent edges and is perhaps the simplest kind of exponential random graph model (of which the Erdős-Rényi model is an easily handled instance), consistency and normality of the maximum likelihood estimator have been established only recently with non-trivial efforts [11, 49, 58], while goodness-of-fit testing still remains an open problem [29]. For general random graph models which do not assume independent edges, very little was known about their asymptotics (but see [3] and [30]) until the groundbreaking work of Chatterjee and Diaconis [10]. By combining the recent theory of graphons with a large deviation result about the Erdős-Rényi model established in [12], they showed that the limiting properties of many exponential random graph models can be obtained by solving a certain variational problem in the graphon space (see Section 2.2 for a summary of this important contribution). Such a result has provided a principled way of resolving the large sample behavior of exponential random graph models and has in fact led to many other novel and insightful results about this class of models. Still, the variational technique in [10] often does not admit an explicit solution and additional work is required.

In this article we further advance our understating of the asymptotics of exponential random graph models by giving a complete characterization of the asymptotic extremal properties of a simple yet challenging 2-parameter exponential random graph model. In more detail, for  $n \geq 2$ , let  $\mathcal{G}_n$  denote the set of all labeled simple (i.e., undirected, with no loops or multiple edges) graphs on  $n$  nodes. Notice that  $|\mathcal{G}_n| = 2^n$ . For a graph  $G_n \in \mathcal{G}_n$  with  $V(G_n)$  vertices and a graph  $H$  with  $V(H)$  vertices, the density homomorphism of  $H$  in  $G_n$  is

$$t(H, G_n) = \frac{|\text{hom}(H, G_n)|}{|V(G_n)|^{|V(H)|}}, \quad (1.1)$$

where  $|\text{hom}(H, G_n)|$  denotes the number of homomorphisms from  $H$  to  $G_n$ , i.e., edge preserving maps from  $V(H)$  to  $V(G_n)$ . Thus,  $t(H, G_n)$  is just the probability that any mapping from  $V(H)$  into  $V(G_n)$  is edge-preserving.

For each  $n$ , we will study the exponential family  $\{\mathbb{P}_\beta, \beta \in \mathbb{R}^2\}$  of probability distribution on  $\mathcal{G}_n$  which assigns to a graph  $G_n \in \mathcal{G}_n$  the probability

$$\mathbb{P}_\beta(G_n) = \exp(n^2(\beta_1 t(H_1, G_n) + \beta_2 t(H_2, G_n) - \psi_n(\beta))), \quad \beta = (\beta_1, \beta_2) \in \mathbb{R}^2, \quad (1.2)$$

where  $H_1 = K_2$  is a single edge,  $H_2$  is a pre-chosen finite simple graph (say a triangle, a two-star, etc.), and  $\psi_n(\beta)$  is the normalizing constant satisfying

$$\exp(n^2 \psi_n(\beta)) = \sum_{G_n \in \mathcal{G}_n} \exp(n^2(\beta_1 t(H_1, G_n) + \beta_2 t(H_2, G_n))). \quad (1.3)$$

Although seemingly simple, this model is well known to exhibit a wealth of non-trivial features (see, e.g., [31, 48]) and challenging asymptotics (see [10]). The two parameters  $\beta_1$  and  $\beta_2$  allow one to adjust the relative influence of different local features – in our case, the edge density and, more importantly, the density of the subgraph  $H_2$ . In statistical physics, we commonly refer to the density of edges as the particle density and the density of subgraph  $H_2$  as the energy density. Positive  $\beta_2$  thus represents a high energy density, i.e., a probability distribution that assigns comparatively larger mass to graphs with high homomorphism density of  $H_2$ , while negative  $\beta_2$  is associated with a low energy density, whereby graphs with a large number of  $H_2$  subgraphs have a smaller likelihood of being observed. Correspondingly we say that the model (1.2) is “attractive” if the parameter  $\beta_2$  is positive, and we say that it is “repulsive” if the parameter  $\beta_2$  is negative. When  $\beta_2$  is zero, the 2-parameter exponential model reduces to the well-studied Erdős-Rényi random graph.

When studying the model (1.2), a natural question to ask is how different values of the parameters  $\beta_1$  and  $\beta_2$  would impact the global structure of a typical random graph  $G_n$  drawn from this model for large  $n$ . Chatterjee and Diaconis [10] showed that, as  $n \rightarrow \infty$ , when  $\beta_2$  is positive or negative but with small magnitude, a random graph  $G_n$  from (1.2) would “look like” an Erdős-Rényi random graph with high probability (in a sense made precise in Section 2.2). They further demonstrated that, as  $n \rightarrow \infty$  and  $\beta_2 \rightarrow -\infty$ , the model will begin to exhibit a peculiar extremal behavior, in the sense that a typical random graph from such a distribution will be close to a random subgraph of a complete multi-partite graph with the number of classes depending on the chromatic number of  $H_2$  (see Section 3.2 for the exact statement of this result).

Below we will generalize the extremal results of [10] and complete an exhaustive study of all the extremal properties of the exponential random graph (1.2) when  $H_2 = K_3$ , i.e., when  $H_2$  is a triangle, which we will refer to as the edge-triangle model. Identifying the extremal properties of the edge-triangle model is not only interesting from a mathematical point of view, but it also gives broad and fundamental insights into the expressive power of the model itself. In our framework, which is an extension of the one in [10], we will simultaneously let the size of the network  $n$  grow unbounded and the natural parameters  $(\beta_1, \beta_2)$  diverge along generic *straight* lines. In our analysis we will elucidate the relationship between all possible directions along which the natural parameters can diverge and the way the model tends to place most of its mass on graph configurations that resemble complete multi-partite graphs for large enough  $n$ . As it turns out, looking just at straight lines is precisely what is needed to categorize all extremal behaviors of the model. We summarize our contributions as follows.

- We will extend the variational analysis technique of [10] to show that the limiting set of the boundary of all extremal distributions of the edge-triangle model in our double asymptotic framework consists of degenerate distributions on all infinite dimensional Turán graphs. We will further exhibit a partition of all the possible half-lines or directions in  $\mathbb{R}^2$  in the form of a collection of cones with apexes at the origin and disjoint interiors, so two sequences of natural parameters diverging along different half-lines in the same cone will yield the same asymptotic extremal behavior. We refer to this result as an asymptotic quantization of the parameter space of the edge-triangle model. We will finally identify a distinguished countable set of critical directions along which the extremal behaviors of the edge-triangle model cannot be uniquely resolved.
- In the second part of the paper, we will present a completely different technique of analysis that relies on the notion of closure of exponential families. Within this second approach, the extremal properties of the model correspond to the asymptotic boundary of the model in the total variation topology. The main advantage of this characterization is its ability to resolve the model also along critical directions. Specifically, we will demonstrate that, along each such direction, the model undergoes, asymptotically, an unexpected phase transition, which we characterize as an asymptotic discontinuity in the natural parametrization. In contrast, in non-asymptotic regimes the natural parametrization is always continuous, even on the boundary of the total variation closure of the model.
- A central ingredient of our analysis is the use of simple yet effective geometric arguments. Both the quantization of the parameter space and the identification of critical directions stem from the dual geometric property of a bounded convex polygon with infinitely many edges, which can be thought of as an asymptotic mean value parametrization of

the edge-triangle exponential model. We expect this framework to apply more generally to other exponential random graph models.

The rest of this paper is organized as follows. In Section 2 we provide some basics of graph limit theory, summarize the main results of [10] and introduce key geometric quantities. In Section 3.1 we investigate the asymptotic behavior of “attractive” 2-parameter exponential random graph models along general straight lines. In Section 3.2 we analyze the asymptotic structure of “repulsive” 2-parameter exponential random graph models along vertical lines. In Sections 3.3 and 4 we examine the asymptotic feature of the edge-triangle model along general straight lines. Finally, Section 5 is devoted to concluding remarks. All the proofs are in the appendix.

## 2. BACKGROUND

Below we will provide some background on the theory of graph limits and its use in exponential random graph models, focusing in particular on the edge-triangle model.

**2.1. Graph limit theory.** A series of recent and fundamental contributions by mathematicians and physicists have led to a unified and elegant theory of limits of sequences of dense graphs, or graphons. See, e.g., [6, 7, 8, 38, 41] and the book [39] for a comprehensive account and references. See also the related work on exchangeable arrays, where some of those results had already been derived: [1, 15, 33, 35, 37]. The graphon framework has provided a new set of tools for representing and studying the asymptotic behavior of graphs, and has become the object of intense research in multiple fields, such as discrete mathematics, statistical mechanics and probability.

Here are the basics of this beautiful theory. According to the original definition of [41], a sequence  $\{G_n\}_{n=1,2,\dots}$  of graphs, where we assume  $G_n \in \mathcal{G}_n$  for each  $n$ , is said to converge when, for every simple graph  $H$ , the limit as  $n \rightarrow \infty$  of the sequences of density homomorphisms of  $H$  into  $G_n$  exists, i.e., when

$$\lim_{n \rightarrow \infty} t(H, G_n) = t(H)$$

for some  $t(H)$ . The main result in [41] is a complete characterization of all limits of converging graph sequences, which are shown to correspond to the functional space  $\mathcal{W}$  of all symmetric measurable functions from  $[0, 1]^2$  into  $[0, 1]$ , called *graph limits* or *graphons*. Specifically, the graph sequence  $\{G_n\}_{n=1,2,\dots}$  converges if and only if there exists one graphon  $f \in \mathcal{W}$  such that, for every simple graph  $H$  with vertex set  $\{1, \dots, k\}$  and edge set  $E(H)$ ,

$$\lim_{n \rightarrow \infty} t(H, G_n) = t(H, f) := \int_{[0,1]^k} \prod_{\{i,j\} \in E(H)} f(x_i, x_j) dx_1 \cdots dx_k. \quad (2.1)$$

Intuitively, we may think of each graphon  $f$  as arising in the following manner: the interval  $[0, 1]$  can be thought of as a continuum of vertices, and  $f(x, y) = f(y, x)$  is the probability that  $x$  and  $y$  are connected by an edge. Indeed, any finite graph  $G_n$  has a natural representation as a graphon of the form

$$f^{G_n}(x, y) = \begin{cases} 1, & \text{if } (\lceil nx \rceil, \lceil ny \rceil) \text{ is an edge in } G_n, \\ 0, & \text{otherwise.} \end{cases} \quad (2.2)$$

In particular, for each finite simple graph  $H$ ,  $t(H, G_n) = t(H, f^{G_n})$  by definition. The theory of graph limits also covers sequences of random graphs, where the notions of convergence in probability and of weak convergence are well defined, as shown in [15].

Among the main advantages of the graphon framework is its ability to represent the limiting properties of sequences of graphs, which are discrete objects, with an abstract functional space that is not only highly interpretable but also enjoys nice analytic properties. For example, for a sequence of Erdős-Rényi random graphs with constant edge probability  $p$ , the limiting graphon is the function that is identically equal to  $p$  on  $[0, 1]^2$ . Similarly, the sequence of Turán graph with  $k$  classes converges to a graphon that takes the value 1 everywhere except along  $k$  diagonal squares, where it is zero (see (2.12)). More importantly perhaps, Lovász and Szegedy [41] have shown that the space of graphons can be endowed with a particular metric, called the *cut metric*, such that convergence of subgraph densities is equivalent to convergence in the cut metric. Formally, for any two graphons  $f$  and  $g$ , define on  $\mathcal{W}$  the cut distance as

$$d_{\square}(f, g) = \sup_{S, T \subseteq [0, 1]} \left| \int_{S \times T} (f(x, y) - g(x, y)) dx dy \right|,$$

where the supremum is over all measurable subsets  $S$  and  $T$ . A non-trivial technical complication is the fact that the topology induced by the cut metric is well defined only up to measure preserving transformations (and up to sets of Lebesgue measure zero), which in the context of finite graphs can be thought of as vertex relabeling. The solution is to work instead on an appropriate quotient space  $\widetilde{\mathcal{W}}$ .

To that end, define on  $\mathcal{W}$  the equivalence relation  $\sim$ , so that  $f \sim g$  if  $f(x, y) = g_{\sigma}(x, y) := g(\sigma x, \sigma y)$  for some measure preserving bijection  $\sigma$  of  $[0, 1]$ . Let  $\tilde{g}$  denote the closure of the orbit  $\{g_{\sigma}\}$  of  $g$  in  $(\mathcal{W}, d_{\square})$ . Since  $d_{\square}$  is invariant under  $\sigma$ , one can then define the natural distance  $\delta_{\square}$  between graphons by  $\delta_{\square}(f, \tilde{g}) = \inf_{\sigma_1, \sigma_2} d_{\square}(f_{\sigma_1}, g_{\sigma_2})$ , where the infimum ranges over all measure preserving bijections  $\sigma_1$  and  $\sigma_2$  (the infimum is always achieved: see Theorem 8.13 in [39]). With some abuse of notation we will also refer to  $\delta_{\square}$  as the cut distance. Finally, the metric space  $(\widetilde{\mathcal{W}}, \delta_{\square})$  is obtained by identifying two graphons  $f$  and  $g$  such that  $\delta_{\square}(f, \tilde{g}) = 0$ . Lovász and Szegedy [41] have showed that convergence of all graph homomorphism densities is equivalent to convergence in  $(\widetilde{\mathcal{W}}, \delta_{\square})$ : a sequence of (possibly random) graph  $\{G_n\}_{n=1, 2, \dots}$  converges to a graphon  $f$  if and only if

$$\lim_n \delta_{\square}(f^{G_n}, f) = 0.$$

where  $f^{G_n}$  is defined in (2.2). The space  $(\widetilde{\mathcal{W}}, \delta_{\square})$  enjoys many nice properties that are essential for the study of exponential random graph models. In particular, it is a compact space and homomorphism densities  $t(H, \cdot)$  are continuous functionals on it.

As a final but important remark, it is worth emphasizing that graphons describe limits of *dense* graphs, i.e., graphs having order  $n^2$  edges. In particular, graphons cannot describe or discern any graph property in the sequence that depends on a number of edges of order  $o(n^2)$ . Therefore, for instance, sequences of graphs with order  $o(n^2)$  edges will trivially converge to the identically zero graphon. By the same token, any sequence of graphs that differ from another sequence converging to the graphon  $f$  by  $o(n^2)$  edges will also converge to the same graphon.

**2.2. Graph limits of exponential random graph models.** In a recent important paper, Diaconis and Chatterjee [10] examined the properties of the metric space  $(\widetilde{\mathcal{W}}, \delta_{\square})$  in order to describe the asymptotic behavior of exponential random graph models. For the purpose of this paper, two results from [10] are particularly significant. The first result, which is an application of a deep large deviation result of [12], is Theorem 3.1 in [10]. When applied to the 2-parameter exponential random graph models mentioned above it implies that the limiting free energy

density  $\psi_\infty(\beta) = \lim_{n \rightarrow \infty} \psi_n(\beta)$  always exists and is given by

$$\psi_\infty(\beta) = \sup_{\tilde{f} \in \tilde{\mathcal{W}}} \left( \beta_1 t(H_1, f) + \beta_2 t(H_2, f) - \iint_{[0,1]^2} I(f) dx dy \right), \quad (2.3)$$

where  $f$  is any representative element of the equivalence class  $\tilde{f}$ , and

$$I(u) = \frac{1}{2} u \log u + \frac{1}{2} (1-u) \log(1-u). \quad (2.4)$$

The second result, Theorem 3.2 in [10], is concerned with the solutions of the above variational optimization problem. In detail, let  $\tilde{F}^*(\beta)$  be the subset of  $\tilde{\mathcal{W}}$  where (2.3) is maximized. Then, the quotient image  $\tilde{f}^{G_n}$  of a random graph  $G_n$  drawn from (1.2) must lie close to  $\tilde{F}^*(\beta)$  with probability vanishing in  $n$ , i.e.,

$$\delta_\square(\tilde{f}^{G_n}, \tilde{F}^*(\beta)) \rightarrow 0 \text{ in probability as } n \rightarrow \infty. \quad (2.5)$$

Due to its complicated structure, the variational problem (2.3) is not always explicitly solvable, and so far major simplification has only been achieved when  $\beta_2$  is positive or negative with small magnitude. It has been discovered in Chatterjee and Diaconis [10] that for  $\beta_2$  lying in these parameter regions,  $G_n$  behaves like an Erdős-Rényi graph  $G(n, u)$  in the large  $n$  limit, where  $u$  is picked randomly from the set  $U$  of maximizers of a reduced form of (2.3):

$$\psi_\infty(\beta) = \sup_{0 \leq u \leq 1} \left( \beta_1 u^{e(H_1)} + \beta_2 u^{e(H_2)} - I(u) \right), \quad (2.6)$$

where  $e(H_i)$  is the number of edges in  $H_i$ . (There are also related results in Häggström and Jonasson [30] and Bhamidi et al. [3].) Some features of the exponential random graph models that remain unexplored are their extremal asymptotic properties, i.e., their large sample behaviors when the natural parameters are large in magnitude. For the 2-parameter exponential random graph models, Chatterjee and Diaconis [10] studied the case in which  $H_1 = K_2$  and  $H_2$  is arbitrary,  $\beta_1$  is fixed and  $\beta_2 \rightarrow -\infty$ . In this paper we will further assume that  $H_2 = K_3$  but make no restrictions on how the natural parameters diverge.

**2.3. The edge-triangle exponential random graph model and its asymptotic geometry.** We focus almost exclusively on the edge-triangle model, which is the exponential random graph model obtained by setting in (1.2)  $H_1 = K_2$  and  $H_2 = K_3$ . Explicitly, in the edge-triangle model the probability of a graph  $G_n \in \mathcal{G}_n$  is

$$\mathbb{P}_\beta(G_n) = \exp \left( n^2 (\beta_1 t(K_2, G_n) + \beta_2 t(K_3, G_n) - \psi_n(\beta)) \right), \quad \beta = (\beta_1, \beta_2) \in \mathbb{R}^2, \quad (2.7)$$

where  $\psi_n(\beta)$  is given in (1.3). Below we describe the asymptotic geometry of this model, which underpins much of our analysis.

To start off, for any  $G_n \in \mathcal{G}_n$ , the vector of the densities of graph homomorphisms of  $K_2$  and  $K_3$  in  $G_n$  takes the form

$$t(G_n) = \begin{pmatrix} t(K_2, G_n) \\ t(K_3, G_n) \end{pmatrix} = \begin{pmatrix} \frac{2E(G_n)}{n^2} \\ \frac{6T(G_n)}{n^3} \end{pmatrix} \in [0, 1]^2, \quad (2.8)$$

where  $E(G_n)$  and  $T(G_n)$  are the number of subgraphs of  $G_n$  isomorphic to  $K_2$  and  $K_3$ , respectively. Since every finite graph can be represented as a graphon, we can extend  $t$  to a map from  $\mathcal{W}$  into  $[0, 1]^2$  by setting (see (2.1))

$$t(f) = \begin{pmatrix} t(K_2, f) \\ t(K_3, f) \end{pmatrix}, \quad f \in \mathcal{W}. \quad (2.9)$$

As we will see, the asymptotic extremal behaviors of the edge-triangle model can be fully characterized by the geometry of two compact subsets of  $[0, 1]^2$ . The first is the set

$$R = \{t(f), f \in \mathcal{W}\} \quad (2.10)$$

of all realizable values of the edge and triangle density homomorphisms as  $f$  varies over  $\mathcal{W}$ . The second set is the convex hull of  $R$ , i.e.,

$$P = \text{convhull}(R). \quad (2.11)$$

Figures 1 and 2 depict  $R$  and  $P$ , respectively.

To describe the properties of the sets  $R$  and  $P$ , we will introduce some quantities that we will use throughout this paper. For  $k = 0, 1, \dots$ , we set  $v_k = t(f^{K_{k+1}})$ , where  $f^{K_1}$  is the identically zero graphon and, for any integer  $k > 1$ ,

$$f^{K_k}(x, y) = \begin{cases} 1 & \text{if } [xk] \neq [yk], \\ 0 & \text{otherwise,} \end{cases} \quad (x, y) \in [0, 1]^2, \quad (2.12)$$

is the Turán graphon with  $k$  classes, with  $[\cdot]$  denoting the integer part of a real number. Thus,

$$v_k = \left( \frac{k}{k+1}, \frac{k(k-1)}{(k+1)^2} \right), \quad k = 0, 1, 2, \dots, \quad (2.13)$$

The name Turán graphon is due to the easily verified fact that

$$v_k = \lim_{n \rightarrow \infty} v_{k,n}, \quad \text{for each } k = 1, 2, \dots,$$

with  $v_{k,n} = t(T(n, k+1))$ , the homomorphism densities of  $K_2$  and  $K_3$  in  $T(n, k+1)$ , a Turán graph on  $n$  nodes with  $k+1$  classes. Turán graphs are well known to provide the solutions of many extremal dense graph problems (see, e.g., [16]), and turn out to be the extremal graphs for the edge-triangle model as well.

The set  $R$  is a classic and well studied object in asymptotic extremal graph theory, even though the precise shape of its boundary was determined only recently (see, e.g., [5, 23, 28, 40] and the book [39]). Letting  $e$  and  $t$  denote the coordinate corresponding to the edge and triangle density homomorphisms, respectively, the upper boundary curve of  $R$  (see Figure 1), is given by the equation  $t = e^{3/2}$ , and can be derived using the Kruskal-Katona theorem (see Section 16.3 of [39]). The lower boundary curve is trickier. The trivial lower bound of  $t = 0$ , corresponding to the horizontal segment, is attainable at any  $0 \leq e \leq 1/2$  by graphons describing the (possibly asymptotic) edge density of subgraphs of complete bipartite graphs. For  $e \geq 1/2$ , the optimal bound was obtained recently by Razborov [47], who established that, using the flag algebra calculus, for  $(k-1)/k \leq e \leq k/(k+1)$  with  $k \geq 2$ ,

$$t \geq \frac{(k-1) \left( k - 2\sqrt{k(k-e(k+1))} \right) \left( k + \sqrt{k(k-e(k+1))} \right)^2}{k^2(k+1)^2}. \quad (2.14)$$

All the curve segments describing the lower boundary of  $R$  are easily seen to be strictly convex, and the boundary points of those segments are precisely the points  $v_k$ ,  $k = 0, 1, \dots$

The following Lemma 2.1 is a direct consequence of Theorem 16.8 in [39] (see page 287 of the same reference for details).

**Lemma 2.1.** (1)  $R = \text{cl}(\{t(G_n), G_n \in \mathcal{G}_n, n = 1, 2, \dots\})$ .  
 (2) *The extreme points of  $P$  are the points  $\{v_k, k = 0, 1, \dots\}$  and the point  $(1, 1) = \lim_{k \rightarrow \infty} v_k$ .*

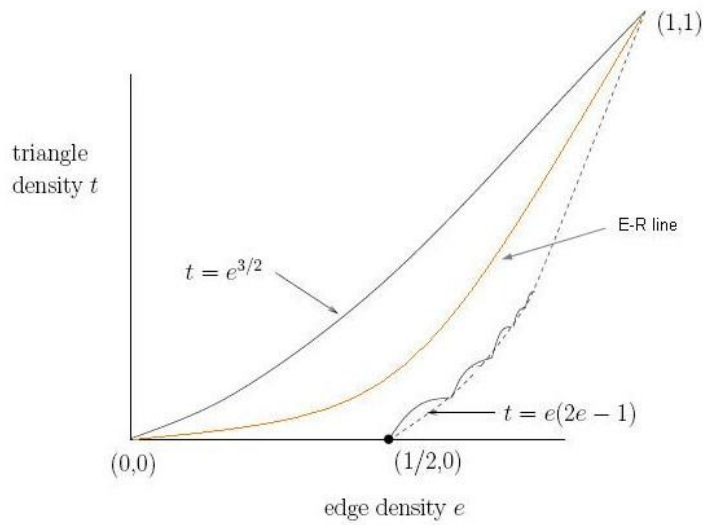


FIGURE 1. The set  $R$  of all feasible edge-triangle homomorphism densities, defined in (2.10).

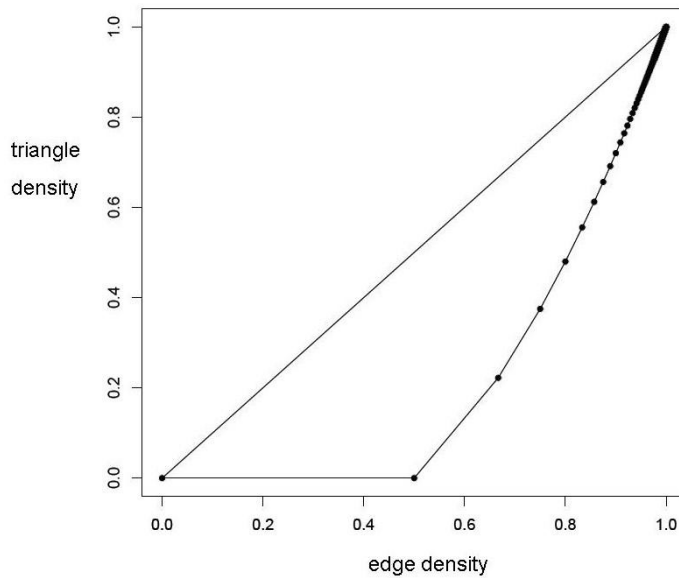


FIGURE 2. The set  $P$  described in (2.11).

The first result of Lemma 2.1 indicates that the set of edge and triangle homomorphism densities of all finite graphs is dense in  $R$ . The second result implies that the boundary of  $P$  consists of infinitely many segments with endpoints  $v_k$ , for  $k = 0, 1, \dots$ , as well as the line segment joining  $v_0 = (0, 0)$  and  $(1, 1) = \lim_{k \rightarrow \infty} v_k$ . In particular,  $P$  is not a polyhedral set.

For  $k = 0, 1, \dots$ , let  $L_k$  be the segment joining the adjacent vertices  $v_k$  and  $v_{k+1}$ , and  $L_{-1}$  the segment joining  $v_0$  and the point  $(1, 1)$ . Each such  $L_k$  is an exposed face of  $P$  of maximal



dimension 1, i.e., a facet. Notice that the length of the segment  $L_k$  decreases monotonically to zero as  $k$  gets larger. For any  $k > 0$ , the slope of the line passing through  $L_k$  is

$$\frac{k(3k+5)}{(k+1)(k+2)},$$

which increases monotonically to 3 as  $k \rightarrow \infty$ .

Simple algebra yields that the facet  $L_k$  is exposed by the vector

$$o_k = \begin{cases} (-1, 1) & \text{if } k = -1. \\ (0, -1) & \text{if } k = 0 \\ \left(1, -\frac{(k+1)(k+2)}{k(3k+5)}\right) & \text{if } k = 1, 2, \dots \end{cases} \quad (2.15)$$

The vectors  $o_k$  will play a key role in determining the asymptotic behavior of the edge-triangle model, so much so that they deserve their own name.

**Definition 2.2.** *The vectors  $\{o_k, k = -1, 0, 1, \dots\}$  are the critical directions of the edge-triangle model.*

It follows that the outer normals to the facets of  $P$  are given by

$$\text{cone}(o_k), \quad k = -1, 0, 1, 2, \dots,$$

i.e., by rays emanating from the origin and going along the vectors  $o_k$ . Finally, for  $k = 0, 1, 2, \dots$  let  $C_k = \text{cone}(o_{k-1}, o_k)$  denote the normal cone to  $P$  at  $v_k$ , a 2-dimensional pointed cone spanned by  $o_{k-1}$  and  $o_k$ . Then, since  $P$  is bounded, for any non-zero  $x \in \mathbb{R}^2$ , there exists one  $k$  for which either  $x \in \text{cone}(o_k)$  or  $x \in C_k^\circ$ .

The normal cones to the faces of  $P$  form a locally finite polyhedral complex of cones, shown in Figure 3. As our results will demonstrate, each one of those cones uniquely identifies one of infinitely many asymptotic extremal behaviors of the edge-triangle model.

### 3. VARIATIONAL ANALYSIS

In this section we characterize the extremal properties of 2-parameter exponential random graphs and especially of the edge-triangle model using the variational approach described in Section 2.2. Chatterjee and Diaconis [10] showed that a typical graph drawn from a 2-parameter exponential random graph model with  $H_1$  an edge and  $H_2$  a fixed graph with chromatic number  $\chi$  is a  $(\chi - 1)$ -equipartite graph when  $n$  is large,  $\beta_1$  is fixed, and  $\beta_2$  is large and negative, i.e., when the two parameters trace a vertical line downward.

In the hope of discovering other interesting extremal behaviors, we investigate the asymptotic structure of 2-parameter exponential random graph models along general straight lines. In particular, we will study sequences of model parameters of the form  $\beta_1 = a\beta_2 + b$ , where  $a$  and  $b$  are constants and  $|\beta_2| \rightarrow \infty$ . Thus, for any  $\beta = (\beta_1, \beta_2) \in \mathbb{R}^2$ , we can regard the quantities  $\tilde{F}^*(\beta)$  and  $\psi_\infty(\beta)$ , defined in Section 2.2, as functions of  $\beta_2$  only, and therefore will write them as  $\tilde{F}^*(\beta_2)$  and  $\psi_\infty(\beta_2)$  instead.

While we only give partial results for general exponential random graphs, we are able to provide a nearly complete characterization of the edge-triangle model. Even more refined results are possible, as will be shown in Section 4.

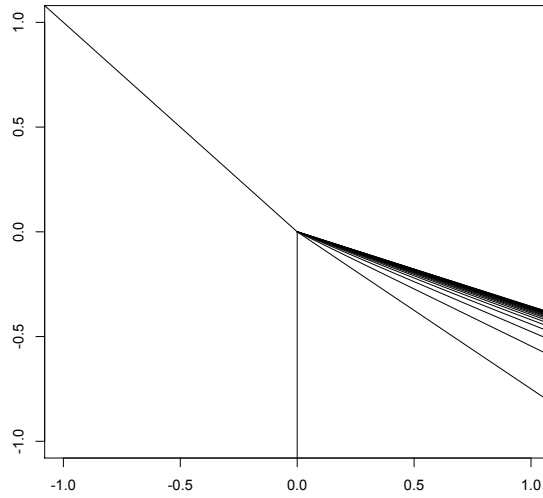


FIGURE 3. Cones, i.e. rays emanating from the origin, generated by the critical directions  $o_{-1}$ ,  $o_0$  and  $o_k$ , for  $k = 1, \dots, 40$ . The plot also provides an accurate depiction of the locally finite polyhedral complex comprised by the normal cones to the faces of the set  $P$  defined in (2.11).

**3.1. Asymptotic behavior of attractive 2-parameter exponential random graph models along general lines.** We will first consider the asymptotic behavior of “attractive” 2-parameter exponential random graph models as  $\beta_2 \rightarrow \infty$ . We will show that for  $H_1$  an edge and  $H_2$  any other finite simple graph, in the large  $n$  limit, a typical graph drawn from the exponential model becomes complete under the topology induced by the cut distance if  $a > -1$  or  $a = -1$  and  $b > 0$ ; it becomes empty if  $a < -1$  or  $a = -1$  and  $b < 0$ ; while for  $a = -1$  and  $b = 0$ , it either looks like a complete graph or an empty graph. Below, for a non-negative constant  $c$ , we will write  $u = c$  when  $u$  is the constant graphon with value  $c$ .

**Theorem 3.1.** *Consider the 2-parameter exponential random graph model (1.2), with  $H_1 = K_2$  and  $H_2$  a different, arbitrary graph. Let  $\beta_1 = a\beta_2 + b$ . Then*

$$\lim_{\beta_2 \rightarrow \infty} \sup_{\tilde{f} \in \tilde{F}^*(\beta_2)} \delta_{\square}(\tilde{f}, \tilde{U}) = 0, \quad (3.1)$$

where the set  $U \subset \mathcal{W}$  is determined as follows:

- $U = \{1\}$  if  $a > -1$  or  $a = -1$  and  $b > 0$ ,
- $U = \{0, 1\}$  if  $a = -1$  and  $b = 0$ , and
- $U = \{0\}$  if  $a < -1$  or  $a = -1$  and  $b < 0$ .

When  $a = -1$  and  $b = 0$ , the limit points of the solution set of the variational problem (2.6) consist of two radically different graphons, one specifying an asymptotic edge density of 1 and the other of 0. This intriguing behavior was captured in [46], where it was shown that there is a continuous curve that asymptotically approaches the line  $\beta_1 = -\beta_2$ , across which the graph transitions from being very sparse to very dense. Unfortunately, the variational technique used in the proof of the theorem does not seem to yield a way of deciding whether only one or both solutions can actually be realized. As we will see next, a similar issue arises when analyzing the asymptotic extremal behavior of the edge-triangle model along critical directions (see Theorem

3.3). In Section 4 we describe a different method of analysis that will allow us to resolve this rather subtle ambiguity within the edge-triangle model and reveal an asymptotic phase transition phenomenon. In particular, Theorem 4.4 there can be easily adapted to provide an analogous resolution of the case  $a = -1$  and  $b = 0$  in Theorem 3.1.

We remark that, using the same arguments, it is also possible to handle the case in which  $\beta_2$  is fixed and  $\beta_1$  diverges along horizontal lines. Then, we obtain the intuitively clear result that, in the large  $n$  limit, a typical random graph drawn from this model becomes complete if  $\beta_1 \rightarrow \infty$ , and empty if  $\beta_1 \rightarrow -\infty$ . We omit the easy proof.

The next two sections deal with the more challenging analysis of the asymptotic behavior of “repulsive” 2-parameter exponential models as  $\beta_2 \rightarrow -\infty$ . As mentioned in the introduction, the asymptotic properties of such models are largely unknown in this region.

**3.2. Asymptotic behavior of repulsive 2-parameter exponential random graph models along vertical lines.** The purpose of this section is to give an alternate and simpler proof of Theorem 7.1 in [10] that uses classic results in extremal graph theory. In addition, this general result covers the asymptotic extremal behavior of the edge-triangle model along the vertical critical direction.

Recall that  $\beta_1$  is fixed and we are interested in the asymptotics of  $\tilde{F}^*(\beta_2)$  and  $\psi_\infty(\beta_2)$  as  $\beta_2 \rightarrow -\infty$ . We point out here that the limit process in  $\beta_2$  may also be interpreted by taking  $\beta_1 = a\beta_2 + b$  with  $a = 0$  and  $b$  large negative. The importance of this latter interpretation will become clear in the next section. Our work here is inspired by related results of Fadnavis and Radin and Sadun [45] in the case of  $H_2$  being a triangle.

**Theorem 3.2** (Chatterjee-Diaconis). *Consider the 2-parameter exponential random graph model (1.2), with  $H_1 = K_2$  and  $H_2$  a different, arbitrary graph. Fix  $\beta_1$ . Let  $r = \chi(H_2)$  be the chromatic number of  $H_2$ . Define*

$$g(x, y) = \begin{cases} 1 & \text{if } [(r-1)x] \neq [(r-1)y]; \\ 0 & \text{otherwise,} \end{cases} \quad (3.2)$$

where  $[\cdot]$  denotes the integer part of a real number. Let  $p = e^{2\beta_1}/(1 + e^{2\beta_1})$ . Then

$$\lim_{\beta_2 \rightarrow -\infty} \sup_{\tilde{f} \in \tilde{F}^*(\beta_2)} \delta_\square(\tilde{f}, \tilde{U}) = 0, \quad (3.3)$$

where the set  $U \subset \mathcal{W}$  is given by  $U = \{pg\}$ .

As explained in [10], the above result can be interpreted as follows: if  $\beta_2$  is negative and large in magnitude and  $n$  is big, then a typical graph  $G_n$  drawn from the 2-parameter exponential model (1.2) looks roughly like a complete  $(\chi(H_2) - 1)$ -equipartite graph with  $1 - p$  fraction of edges randomly deleted, where  $p = e^{2\beta_1}/(1 + e^{2\beta_1})$ .

**3.3. Asymptotic quantization of edge-triangle model along general lines.** In this section we conduct a thorough analysis of the asymptotic behavior of the edge-triangle model as  $\beta_2 \rightarrow -\infty$ . As usual, we take  $\beta_1 = a\beta_2 + b$ , where  $a$  and  $b$  are fixed constants. The  $a = 0$  situation is a special case of what has been discussed in the previous section: If  $n$  is large, then a typical graph  $G_n$  drawn from the edge-triangle model looks roughly like a complete bipartite graph with  $1/(1 + e^{2b})$  fraction of edges randomly deleted. It is not too difficult to establish that if  $a > 0$ , then independent of  $b$ , a typical graph  $G_n$  becomes empty in the large  $n$  limit. Intuitively, this should be clear:  $\beta_1$  and  $\beta_2$  both large and negative entail that  $G_n$  would have minimal edge and triangle densities. However, the case  $a < 0$  leads to an array of non-trivial and intriguing extremal behaviors for the edge-triangle model, and they are described in our

next result. We emphasize that our analysis relies in a fundamental way on the explicit characterization by Razborov [47] of the lower boundary of the set  $R$  of (the closure of) all edge and triangle density homomorphisms (see (2.14)) and on the fact that the extreme points of  $P$  are the points  $\{v_k, k = 0, 1, \dots\}$ , given in (2.13). Recall that these points correspond to the density homomorphisms of the Turán graphons  $f^{K_{k+1}}$ ,  $k = 0, 1, \dots$ , as shown in (2.12).

**Theorem 3.3.** *Consider the edge-triangle exponential random graph model (2.7). Let  $\beta_1 = a\beta_2 + b$  with  $a < 0$  and, for  $k \geq 0$ , let  $a_k = -\frac{k(3k+5)}{(k+1)(k+2)}$ . Then,*

$$\lim_{\beta_2 \rightarrow -\infty} \sup_{\tilde{f} \in \tilde{F}^*(\beta_2)} \delta_{\square}(\tilde{f}, \tilde{U}) = 0, \quad (3.4)$$

where the set  $U \subset \mathcal{W}$  is determined as follows:

- $U = \{f^{K_{k+2}}\}$  if  $a_k > a > a_{k+1}$  or  $a = a_k$  and  $b > 0$ ,
- $U = \{f^{K_{k+1}}, f^{K_{k+2}}\}$  if  $a = a_k$  and  $b = 0$ , and
- $U = \{f^{K_{k+1}}\}$  if  $a = a_k$  and  $b < 0$ .

**Remark.** *Notice that the case  $a = a_k$  and  $b = 0$  corresponds to the critical direction  $o_k$ , for  $k = 1, 2, \dots$*

The above result says that, if  $\beta_1 = a\beta_2 + b$  with  $a < 0$  and  $\beta_2$  large negative, then in the large  $n$  limit, any graph drawn from the edge-triangle model is indistinguishable in the cut metric topology from a complete  $(k+2)$ -equipartite graph if  $a_k > a > a_{k+1}$  or  $a = a_k$  and  $b > 0$ ; it looks like a complete  $(k+1)$ -equipartite graph if  $a = a_k$  and  $b < 0$ ; and for  $a = a_k$  and  $b = 0$ , it either behaves like a complete  $(k+1)$ -equipartite graph or a complete  $(k+2)$ -equipartite graph. Lastly it becomes complete if  $a \leq \lim_{k \rightarrow \infty} a_k = -3$ . Overall, these results describe in a precise manner the array of all possible asymptotic extremal behaviors of the edge-triangle model, and link them directly to the geometry of the natural parameter space as captured by the polyhedral complex of cones shown in Figure 3.

When  $a = a_k$  and  $b = 0$ , for any  $k = 0, 1, \dots$ , i.e., when the parameters diverge along the critical direction  $o_k$ , Theorem 3.3 suffers from the same ambiguity as Theorem 3.1: the limit points of the solution set of the variational problem (2.3) as  $\beta_2 \rightarrow -\infty$  are Turán graphons with  $k+1$  and  $k+2$  classes. Though already quite informative, this result remains somewhat unsatisfactory because it does not indicate whether both such graphons are actually realizable in the limit and in what manner. As we remarked in the discussion following Theorem 3.1, our method of proof, largely based on and inspired by the results in [10], does not seem to suggest a way of clarifying this issue. In the next section we will consider an alternative asymptotic framework yielding different types of converge guarantees that will not only conform to the results of Theorem 3.3, but will in fact resolve the aforementioned ambiguities observed along critical directions.

### 3.4. Probabilistic convergence of sequence of graphs from the edge-triangle model.

The results obtained in Sections 3.1, 3.2 and 3.3 characterize the extremal asymptotic behavior of the edge-triangle model through functional convergence in the cut topology within the space  $\tilde{\mathcal{W}}$ . Our explanation of such results though has more of a probabilistic flavor. Here we briefly show how this interpretation is justified. By combining (2.5), established in Theorem 3.2 of [10], with the theorems in Sections 3.1, 3.2 and 3.3, and a standard diagonal argument, we can deduce the existence of subsequences of the form  $\{(n_i, \beta_{2,i})\}_{i=1,2,\dots}$ , where  $n_i \rightarrow \infty$  and  $\beta_{2,i} \rightarrow \infty$  or  $-\infty$  as  $i \rightarrow \infty$ , such that the following holds. For fixed  $a$  and  $b$ , let  $\{G_i\}_{i=1,2,\dots}$  be

a sequence of random graphs drawn from the sequence of edge-triangle models with node sizes  $\{n_i\}$  and parameter values  $\{(a\beta_{2,i} + b, \beta_{2,i})\}$ . Then

$$\delta_{\square}(\tilde{f}^{G_i}, \tilde{U}) \rightarrow 0 \text{ in probability as } i \rightarrow \infty,$$

where the set  $U \subset \mathcal{W}$ , which depends on  $a$  and  $b$ , is described in Theorems 3.1, 3.2 and 3.3. In Section 4.5 we will obtain a very similar result by entirely different means.

#### 4. FINITE $n$ ANALYSIS

In the remainder part of this paper we will present an alternative analysis of the asymptotic behavior of the edge-triangle model using directly the properties of the exponential families and their closure in the finite  $n$  case instead of the variational approach of [10, 45, 46]. Though the results in this section are seemingly similar to the ones in Section 3, we point out that there are marked differences. First, while in Section 3 we study convergence in the cut metric for the quotient space  $\widetilde{\mathcal{W}}$ , here we are concerned instead with convergence in total variation of the edge and triangle homomorphism densities. Secondly, the double asymptotics, in  $n$  and in the magnitude of  $\beta$ , are not the same. In Section 3, the system size  $n$  goes to infinity first followed by the divergence of the parameter  $\beta_2$  to positive or negative infinity. In contrast, here we first let the magnitude of the natural parameter  $\beta$  diverge to infinity so as to isolate a simpler “restricted” edge-triangle, and then study its limiting properties as  $n$  grows. Though both approaches are in fact asymptotic, we characterize the latter as “finite  $n$ ”, to highlight the fact that we are not working with a limiting system and because, even with finite  $n$ , the extremal properties already begin to emerge. Despite these differences, the conclusions we can derive from both types of analysis are rather similar. Furthermore, they imply a nearly identical convergence in probability in the cut topology (see Sections 3.4 and 4.5).

Besides giving a rather strong form of asymptotic convergence, one of the appeals of the finite  $n$  analysis consists in its ability to provide a more detailed categorization of the limiting behavior of the model along critical directions using simple geometric arguments based on the dual geometry of  $P$ , the convex hull of edge-triangle homomorphism densities. Specifically, we will demonstrate that, asymptotically, the edge-triangle model undergoes phase transitions along critical directions, where its homomorphism densities will converge in total variation to the densities of one of two Turán graphons, both of which are realizable. In addition we are able to state precise conditions on the natural parameters for such transitions to occur.

**4.1. Exponential families.** We begin by reviewing some of the standard theory of exponential families and their closure in the context of the edge-triangle model. We refer the readers to Barndorff-Nielsen [2] and Brown [9] for exhaustive treatments of exponential families, and to Csiszár and Matúš [13, 14], Geyer [25], and Rinaldo et al. [48] for specialized results on the closure of exponential families directly relevant to our problem.

Recall that we are interested in the exponential family of probability distributions on  $\mathcal{G}_n$  such that, for a given choice of the natural parameters  $\beta \in \mathbb{R}^2$ , the probability of observing a network  $G_n \in \mathcal{G}_n$  is

$$\mathbb{P}_{\beta}(G_n) = \exp(n^2 (\langle \beta, t(G_n) \rangle - \psi_n(\beta))), \quad \beta \in \mathbb{R}^2, \quad (4.1)$$

where  $\psi_n(\beta)$  is the normalizing constant and the function  $t(\cdot)$  is given in (2.8). We remark that the above model assigns the same probability to all graphs in  $\mathcal{G}_n$  that have the same image under  $t(\cdot)$ . We let  $S_n = \{t(G_n), G_n \in \mathcal{G}_n\} \subset [0, 1]^2$  be the set of all possible vectors of densities of graph homomorphisms of  $K_2$  and  $K_3$  over the set  $\mathcal{G}_n$  of all simple graphs on  $n$  nodes (see (2.8)). By sufficiency (see, e.g., [9]), the family on  $\mathcal{G}_n$  will induce the exponential family of

probability distributions  $\mathcal{E}_n = \{\mathbb{P}_{n,\beta}, \beta \in \mathbb{R}^2\}$  on  $S_n$ , such that the probability of observing a point  $x \in S_n$  is

$$\mathbb{P}_{n,\beta}(x) = \exp(n^2(\langle \beta, x \rangle - \psi_n(\beta))) \nu_n(x), \quad \beta \in \mathbb{R}^2. \quad (4.2)$$

where  $\nu_n(x) = |\{t^{-1}(x)\}|$  is the measure on  $S_n$  induced by the counting measure on  $\mathcal{G}_n$  and  $t(\cdot)$ . For each  $n$ , the family  $\mathcal{E}_n$  has finite support not contained in any lower dimensional set (see Lemma 4.1 below) and, therefore, is full and regular and, in particular, steep.

We will study the limiting behavior of sequences of models of the form  $\{\mathbb{P}_{n,\beta+ro}\}$ , where  $\beta$  and  $o$  are fixed vectors in  $\mathbb{R}^2$  and  $n$  and  $r$  are parameters both tending to infinity. The interpretation of the parameter  $n$  is immediate: it specifies the size of the network. While it may be tempting to regard  $n$  as a surrogate for an increasing sample size, this would in fact be incorrect. Models parametrized by different values of  $n$  and the same  $r$  cannot be embedded in any sequence of consistent probability measures, since the probability distribution corresponding to the smaller network cannot be obtained from the other by marginalization. See [52] for details. As for the other parameter  $r$ , it quantifies the rate at which the norm of the natural parameters  $\beta + ro$  diverges to infinity along the direction  $o$ . We will show that different choices of the direction  $o$  will yield different extremal behaviors of the model and we will categorize the variety of those behaviors as a function of  $o$  and, whenever it matters, of  $\beta$ . A key feature of our analysis is the direct link to the geometric properties of the polyhedral complex  $\{C_k, k = -1, 0, \dots\}$  defined by the set  $P$  (see Section 2.3).

Overall, the results of this section are obtained with non-trivial extensions of techniques described in the exponential families literature. Indeed, for fixed  $n$ , determining the limiting behaviors of the family  $\mathcal{E}_n$  along sequences of natural parameters  $\{\beta + ro\}_{r \rightarrow \infty}$  for each unit norm vector  $o$  and each  $\beta$  is the main technical ingredient in computing the total variation closure of  $\mathcal{E}_n$ . In particular Geyer [25] refers to the directions  $o$  as the “directions of recession” of the model. The relevance of the directions of recession to the asymptotic behavior of exponential random graphs is now well known, as demonstrated in the work of Handcock [31] and Rinaldo et al. [48].

**4.2. Finite  $n$  geometry.** As we saw in Section 3, the critical directions are determined by the limiting object  $P$ . For finite  $n$ , an analogous role is played by the convex support of  $\mathcal{E}_n$ , which is defined as

$$P_n = \text{convhull}(S_n) \subset [0, 1]^2.$$

The interior of  $P_n$  is equal to all possible expected values of the sufficient statistics:  $P_n^\circ = \{\mathbb{E}_\beta(t(G_n)), \beta \in \mathbb{R}^2\}$ . Thus, it provides a different parametrization of  $\mathcal{E}_n$ , known as the mean-value parametrization (see, e.g., [2, 9]). Unlike the natural parametrization, the mean value parametrization has explicit geometric properties that turn out to be particularly convenient in order to describe the closure of  $\mathcal{E}_n$ , and, ultimately, the asymptotics of the model.

The next lemma characterizes the geometric properties of  $P_n$ . The most significant of these properties is that  $\lim_n P_n = P$ , an easy result that turns out to be the key for our analysis. Recall that we denote by  $T(n, r)$  any Turán graph on  $n$  nodes with  $r$  classes. For  $k = 0, 1, \dots, n-1$ , set  $v_{k,n} = t(T(n, k+1))$  and let  $L_{k,n}$  denote the line segment joining  $v_{k,n}$  and  $v_{k+1,n}$ .

- Lemma 4.1.** (1) *The polytope  $P_n$  is spanned by the points  $\{v_{k,n}, k = 0, 1, \dots, \lceil n/2 \rceil - 1\}$  and  $v_{n-1,n}$ .*  
 (2)  $\lim_n P_n = P$ .  
 (3) *If  $n$  is a multiple of  $(k+1)(k+2)$ , then  $v_{k,n} = v_k$  and  $v_{k+1,n} = v_{k+1}$ . In addition, for all such  $n$ ,  $L_{k,n} \cap S_n = \{v_k, v_{k+1}\}$ .*

Part 2. of Lemma 4.1 implies that for each  $k$ ,  $\lim_n v_{k,n} = v_k$ , a fact that will be used in Theorem 4.2 to describe the asymptotics of the model along generic (i.e., non-critical directions). This conclusion still holds if the polytopes  $P_n$  are the convex hulls of isomorphism, not homomorphism, densities. In this case, however, we have that  $P_n \supset P$  for each  $n$  (see [18]). The seemingly inconsequential fact stated in part 3. is instead of technical significance for our analysis of the phase transitions along critical directions, as will be described in Theorem 4.3. We take note that when  $n$  is not a multiple of  $(k+1)(k+2)$ , part 3. does not hold in general.

**4.3. Asymptotics along generic directions.** Our first result, which gives similar finding as in Section 3.3 shows that, for large  $n$ , if the distribution is parametrized by a vector with very large norm, then almost all of its mass will concentrate on the isomorphic class of a Turán graph (possibly the empty or complete graph). Which Turán graph it concentrate on will essentially depend on the “direction” of the parameter vector with respect to the origin. Furthermore, there is an array of extremal directions that will give the same isomorphic class of a Turán graph.

**Theorem 4.2.** *Let  $o$  and  $\beta$  be vectors in  $\mathbb{R}^2$  such that  $o \neq o_k$  for  $k = -1, 0, 1, \dots$  and let  $k$  be such that  $o \in C_k^\circ$ . For any  $0 < \epsilon < 1$  arbitrarily small, there exists an  $n_0 = n_0(\beta, \epsilon, o) > 0$  such that the following holds: for every  $n > n_0$ , there exists an  $r_0 = r_0(\beta, \epsilon, o, n) > 0$  such that for all  $r > r_0$ ,*

$$\mathbb{P}_{n, \beta + r o}(v_{k,n}) > 1 - \epsilon.$$

**Remark.** *If in the theorem above we consider only values of  $n$  that are multiples of  $(k+1)(k+2)$  then, by Lemma 4.1,  $v_{k,n} = v_k$  for all such  $n$ , which implies convergence in total variation to the point mass at  $v_k$ .*

The theorem shows that *any* choice of  $o \in C_k^\circ$  will yield the same asymptotic (in  $n$  and  $r$ ) behavior, captured by the Turán graph with  $k+1$  classes. This can be further strengthened to show that the convergence is uniform in  $o$  over compact subsets of  $C_k^\circ$ . See [48] for details. Interestingly, the initial value of  $\beta$  does not play any role in determining the asymptotics of  $\mathbb{P}_{n, \beta + r o}$ , which instead depends solely on which cone  $C_k$  contains in its interior the direction  $o$ . Altogether, Theorem 4.2 can be interpreted as follows: the interiors of the cones of the infinite polyhedral complex  $\{C_k, k = -1, 0, 1, \dots\}$  represent equivalence classes of “extremal directions” of the model, whereby directions in the same class will parametrize, for large  $n$  and  $r$ , the same degenerate distribution on some Turán graph.

**4.4. Asymptotics along critical directions.** Theorem 4.2 provides a complete categorization of the asymptotics (in  $n$  and  $r$ ) of probability distributions of the form  $P_{n, \beta + r o}$  for any generic direction  $o$  other than the critical directions  $\{o_k, k = -1, 0, 1, \dots\}$ . We will now consider the more delicate cases in which  $o = o_k$  for some  $k$ . According to Theorem 3.3, in these instances the typical graph drawn from the model will converge (as  $n$  and  $r$  grow and in the cut metric) to a large Turán graph, whose number of classes is not entirely specified.

Our first result characterizes such behavior along sub-sequences of the form  $n = j(k+1)(k+2)$ , for  $j = 1, 2, \dots$  and  $k$  a positive integer. Interestingly, and in contrast with Theorem 4.2, the limiting behavior along any critical direction  $o_k$  depends on  $\beta$  in a discontinuous manner. Before stating the result we will need to introduce some additional notation. Let  $l_k \in \mathbb{R}^2$  be the unit norm vector spanning the one-dimensional linear subspace  $\mathcal{L}_k$  given by the line through the origin parallel to  $L_k$ , where  $k > 0$ , so  $l_k$  is just a rescaling of the vector

$$\left(1, \frac{k(3k+5)}{(k+1)(k+2)}\right).$$

Next let  $H_k = \{x \in \mathbb{R}^2: \langle x, l_k \rangle = 0\} = \mathcal{L}_k^\perp$  be the line through the origin consisting of the linear subspace orthogonal to  $\mathcal{L}_k$  and let

$$H_k^+ = \{x \in \mathbb{R}^2: \langle x, l_k \rangle > 0\} \quad \text{and} \quad H_k^- = \{x \in \mathbb{R}^2: \langle x, l_k \rangle < 0\} \quad (4.3)$$

be the positive and negative halfspaces cut out by  $H_k$ , respectively. Notice that the linear subspace  $\mathcal{L}_k^\perp$  is spanned by the vector  $o_k$  defined in (2.15).

We will make the simplifying assumption that  $n$  is a multiple of  $(k+1)(k+2)$ . This implies, in particular, that  $v_{k,n} = v_k$  and  $v_{k+1,n} = v_{k+1}$  are both vertices of  $P_n$  and that the line segment  $L_{k,n} = L_k$  is a facet of  $P_n$  whose normal cone is spanned by the point  $o_k$ .

**Theorem 4.3.** *Let  $k$  be a positive integer,  $\beta \in \mathbb{R}^2$  and  $0 < \epsilon < 1$  be arbitrarily small. Then there exists an  $n_0 = n_0(\beta, \epsilon, k) > 0$  such that the following holds: for every  $n > n_0$  and a multiple of  $(k+1)(k+2)$  there exists an  $r_0 = r_0(\beta, \epsilon, k, n) > 0$  such that for all  $r > r_0$ ,*

- if  $\beta \in H_k^+$  or  $\beta \in H_k$  then  $\mathbb{P}_{n, \beta + r o_k}(v_{k+1}) > 1 - \epsilon$ ,
- if  $\beta \in H_k^-$ , then  $\mathbb{P}_{n, \beta + r o_k}(v_k) > 1 - \epsilon$ .

The previous result shows that, for large values of  $r$  and  $n$  (assumed to be a multiple of  $(k+1)(k+2)$ ), the probability distribution  $P_{n, \beta + r o_k}$  will be concentrated almost entirely on either  $v_k$  or  $v_{k+1}$ , depending on which side of  $H_k$  the vector  $\beta$  lies. In particular, the actual value of  $\beta$  does not play any role in the asymptotics: only its position relative to  $H_k$  matters. An interesting consequence of our result is the discontinuity of the natural parametrization along the line  $H_k$  in the limit as both  $n$  and  $r$  tend to infinity. This is in stark contrast to the limiting behavior of the same model when  $n$  is infinity and  $r$  tends to infinity: in this case the natural parametrization is a smooth (though non-minimal) parametrization.

We will now consider the critical directions  $o_{-1}$  and  $o_0$  (see (2.15)), which are not covered by Theorem 4.3. We will first describe the behavior of  $\mathcal{E}_n$  along the direction of recession  $o_{-1,n} := \left(-1, \frac{n}{n-2}\right)$ . This is the outer normal to the segment joining the vertices  $v_{0,n} = (0, 0)$  and  $v_{n-1,n} = \left(1 - \frac{1}{n}, \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\right)$  of  $P_n$ , representing the empty and the complete graph, respectively. Notice that  $o_{-1,n} \rightarrow o_{-1}$  as  $n \rightarrow \infty$ . In this case, we show that, for  $n$  and  $r$  large, the probability  $P_{n, \beta + r o_n}$ , with  $\beta = (\beta_1, \beta_2)$ , assigns almost all of its mass to the empty graph when  $\beta_1 + \beta_2 < 0$  and to the complete graph when  $\beta_1 + \beta_2 > 0$ , and it is uniform over  $v_{0,n}$  and  $v_{n-1,n}$  when  $\beta_1 n(n-1) + \beta_2(n-1)(n-2) = 0$ .

**Theorem 4.4.** *Let  $\beta = (\beta_1, \beta_2) \in \mathbb{R}^2$  be a fixed vector and  $0 < \epsilon < 1$  be arbitrarily small. Then for every  $n$  there exists an  $r_0 = r_0(\beta, \epsilon, n)$  such that, for all  $r > r_0$ , the total variation distance between  $P_{n, \beta + r o_n}$  and the probability distribution which assigns to the points  $v_{0,n}$  and  $v_{n-1,n}$  the probabilities*

$$\frac{1}{1 + \exp(\beta_1 n(n-1) + \beta_2(n-1)(n-2))}$$

and

$$\frac{\exp(\beta_1 n(n-1) + \beta_2(n-1)(n-2))}{1 + \exp(\beta_1 n(n-1) + \beta_2(n-1)(n-2))},$$

respectively, is less than  $\epsilon$ .

In the last result of this section we will turn to the critical direction  $o_0 = (0, -1)$ , which, for every  $n \geq 2$ , is the outer normal to the horizontal facet of  $P_n$  joining the points  $(0, 0)$  and

$$v_{1,n} = \left(\frac{2\lceil n/2 \rceil(n - \lceil n/2 \rceil)}{n^2}, 0\right),$$



which we denote with  $L_{0,n}$ . Let  $\mathcal{G}_{n,0}$  denote the subset of  $\mathcal{G}_n$  consisting of triangle free graphs. For each  $n$ , consider the exponential family  $\{\mathbb{Q}_{n,\beta_1}, \beta_1 \in \mathbb{R}\}$  of probability distributions on  $L_{0,n} \cap S_n$  given by

$$\mathbb{Q}_{n,\beta_1}(x) = \exp(n^2(\beta_1 x_1 - \phi_n(\beta_1))) \nu_n(x), \quad x \in L_{n,0} \cap S_n, \quad \beta_1 \in \mathbb{R}, \quad (4.4)$$

where  $\phi_n(\beta_1)$  is the normalizing constant and  $\nu_n(x) = |\{t^{-1}(x)\}|$  is the measure on  $L_{0,n}$  induced by the counting measure on  $\mathcal{G}_{n,0}$  and  $t(\cdot)$ .

**Theorem 4.5.** *Let  $\beta = (\beta_1, \beta_2) \in \mathbb{R}^2$  be a fixed vector and  $0 < \epsilon < 1$  an arbitrary number. Then for every  $n$  there exists an  $r_0 = r_0(\beta, \epsilon, n)$  such that for all  $r > r_0$  the total variation distance between  $\mathbb{P}_{n,\beta+r_0}$  and  $\mathbb{Q}_{n,\beta_1}$  is less than  $\epsilon$ .*

When compared to Theorem 3.2, Theorem 4.5 is less informative, as the class of triangle free graphs is larger than the class of subgraphs of the Turán graphs  $T(n, 2)$ . We conjecture that this gap can indeed be resolved by showing that, for each  $\beta_1$ ,  $\mathbb{Q}_{n,\beta_1}$  assigns a vanishingly small mass to the set of all triangle free graphs that are not subgraphs of some  $T(n, 2)$  as  $n \rightarrow \infty$ . See [19] for relevant results.

**4.5. From convergence in total variation to stochastic convergence in the cut distance.** The results presented so far in this section concern convergence in total variation of the homomorphism densities of edges and triangles to point mass distributions at points  $v_{k,n}$ . They describe a rather different type of asymptotic guarantees from the one obtained in Section 3, whereby convergence occurs in the functional space  $\mathcal{W}$  under the cut metric. Nonetheless, both sets of results are qualitatively similar and lend themselves to nearly identical interpretations. Here we sketch how the total variation convergence results imply convergence in probability in the cut metric along subsequences. Notice that this is precisely the same type of conclusions we obtained at the end of the variational analysis, as remarked in Section 3.4.

We will let  $n$  be of the form  $j(k+1)(k+2)$  for  $j = 1, 2, \dots$ . For simplicity we consider a direction  $o$  in the interior of  $C_k$ , for some  $k$ . By Theorem 4.2 and using a standard diagonal argument, there exists a subsequence  $\{(n_i, r_i)\}_{i=1,2,\dots}$  such that the sequence of probability measures  $\{\mathbb{P}_{n_i, \beta+r_i o}\}_{i=1,2,\dots}$  converges in total variation to the point mass at  $v_k$ . Thus, for each  $\epsilon > 0$ , there exists an  $i_0 = i_0(\epsilon)$  such that, for all  $i > i_0$ , the probability that a random graph  $G_{n_i}$  drawn from the probability distribution  $\mathbb{P}_{n_i, \beta+r_i o}$  is such that  $t(G_i) \neq v_k$  is less than  $\epsilon$ . More formally, let  $\mathcal{A}_i$  be the event that  $t(G_i) = v_k$  and for notational convenience denote  $\mathbb{P}_{n_i, \beta+r_i o}$  by  $\mathbb{P}_i$ . Thus, for each  $i > i_0$ ,  $\mathbb{P}_i(\mathcal{A}_i) > 1 - \epsilon$ . Let  $H$  be any finite graph. Then, denoting with  $\mathbb{E}_i$  the expectation with respect to  $\mathbb{P}_i$  and with  $1_{\mathcal{A}_i}$  the indicator function of  $\mathcal{A}_i$ , we have

$$\mathbb{E}_i[t(H, G_i)] = \mathbb{E}_i[t(H, G_i)1_{\mathcal{A}_i}] + \mathbb{E}_i[t(H, G_i)1_{\mathcal{A}_i^c}],$$

where

$$\mathbb{E}_i[t(H, G_i)1_{\mathcal{A}_i}] = t(H, T(n_i, k+1)) = t(H, f^{K_{k+1}}),$$

since, given our assumption on the  $n_i$ 's, the point in  $\widetilde{\mathcal{W}}$  corresponding to  $T(n_i, k+1)$  is  $\tilde{f}^{K_{k+1}}$  for all  $i$ . Thus, using the fact that density homomorphisms are bounded by 1,

$$\mathbb{E}_i[t(H, G_i)] - t(H, f^{K_{k+1}}) = \mathbb{E}_i[t(H, G_i)1_{\mathcal{A}_i^c}] \leq \mathbb{P}_i(\mathcal{A}_i^c) = \epsilon$$

for all  $i > i_0$ . Thus, we conclude that  $\lim_i \mathbb{E}_i[t(H, G_i)] = t(H, f^{K_{k+1}})$  for each finite graph  $H$ . By Corollary 3.2 in [15], as  $i \rightarrow \infty$ ,

$$\delta_{\square}(\tilde{f}^{G_i}, \tilde{f}^{K_{k+1}}) \rightarrow 0 \quad \text{in probability.}$$

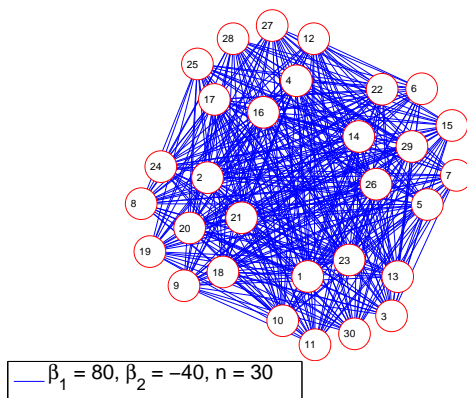


FIGURE 4. A simulated realization of the exponential random graph model on 30 nodes with edges and triangles as sufficient statistics, where the initial value  $\beta = (0, 0)$ ,  $r = 80$ , and the generic direction  $o = (1, -1/2)$  in  $C_3^o$ . The structure of the simulated graph matches the predictions of Theorem 4.2.

Similar arguments apply to the case in which  $o = o_k$  for some  $k > 0$ . Using instead Theorem 4.3, we obtain that, if  $\{G_{n_i}\}_{i=1,2,\dots}$  is a sequence of random graphs drawn from the sequence of probability distributions  $\{\mathbb{P}_{n_i, \beta + r_i o}\}$ , then, as  $i \rightarrow \infty$ ,

$$\delta_{\square}(\tilde{f}^{G_i}, \tilde{f}^{K_{k+2}}) \rightarrow 0, \quad \text{if } \beta \in H_k^+ \text{ or } \beta \in H_k$$

and

$$\delta_{\square}(\tilde{f}^{G_i}, \tilde{f}^{K_{k+1}}) \rightarrow 0, \quad \text{if } \beta \in H_k^-,$$

in probability.

**4.6. Illustrative Figures.** We have validated our theoretical findings with simulations of the edge-triangle model under various specifications on the model parameters. Figure 4 depicts a typical realization from the model when  $n = 30$  and  $o$  is in the interior of  $C_3^o$ . As predicted by Theorem 4.2 the resulting graph is complete equipartite with 4 classes. Figures 5, 6 and 7 exemplify instead the results of Theorem 4.3. For these simulations, we consider the critical direction  $o_1 = (1, -3/4)$  and again a network size of  $n = 30$ , and then vary the initial values of  $\beta$ . Figures 5 and 7 show respectively the outcome of two typical draws when  $\beta$  is in  $H_1^-$  and  $H_1^+$ , respectively. As predicted by our theorem, we obtain a complete bipartite and tripartite graph. Figure 6 depicts instead the case of  $\beta$  exactly along the hyperplane  $H_1$ . As predicted by our theory, a typical realization would be again a complete tripartite graph.

As a final remark, simulating from the extremal parameter configurations we described using off-the-shelf MCMC methods (see, e.g., [26, 34] and, for a convergence result, [10]) can be extremely difficult. This is due to the fact that under these extremal settings, the model places most of its mass on only one or two types of Turán graphs, and the chance of a chain being able to explore adequately the space of graphs using local moves and to eventually reach the configuration of highest energy is essentially minuscule.

## 5. CONCLUDING REMARKS

This paper shows that asymptotically the “repulsive” edge-triangle exponential random graph model exhibits quantized behavior along general straight lines. As we continuously vary the

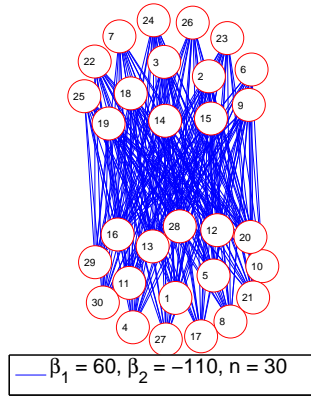


FIGURE 5. A simulated realization of the exponential random graph model on 30 nodes with edges and triangles as sufficient statistics, where the initial value  $\beta = (20, -80)$  in  $H_1^-$ ,  $r = 40$ , and the critical direction  $o_1 = (1, -3/4)$ . The structure of the simulated graph matches the predictions of Theorem 4.3.

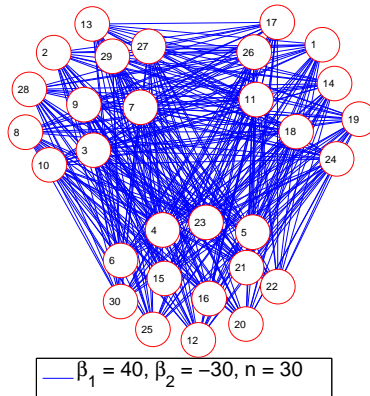


FIGURE 6. A simulated realization of the exponential random graph model on 30 nodes with edges and triangles as sufficient statistics, where the initial value  $\beta = (0, 0)$  in  $H_1$ ,  $r = 40$ , and the critical direction  $o_1 = (1, -3/4)$ . The structure of the simulated graph matches the predictions of Theorem 4.3.

slopes of these lines, a typical graph drawn from this model jumps from one complete multipartite graph to another, and the jumps happen precisely at the normal lines of a polyhedral set with infinitely many facets. There are two limit processes being considered here. First, the number of vertices of the random graph is tending to infinity, since it specifies the size of a large network. Second, the energy parameter is tending to negative infinity, suggesting that the system is near the energy ground state. Our results thus depict a large system that progressively transitions through finer and finer crystal structure at low energy density.

To verify these results, we take two different approaches. The first one is a variational approach as in [10, 45, 46]. In this approach we take the system size  $n$  as infinite and examine the limiting behavior of the model as the energy parameter  $\beta_2$  goes to negative infinity. This

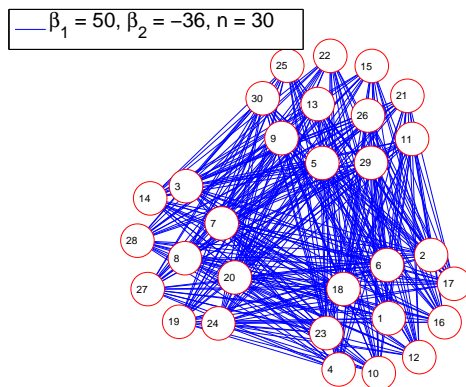


FIGURE 7. A simulated realization of the exponential random graph model on 30 nodes with edges and triangles as sufficient statistics, where the initial value  $\beta = (10, -6)$  in  $H_1^+$ ,  $r = 40$ , and the critical direction  $o_1 = (1, -3/4)$ . The structure of the simulated graph matches the predictions of Theorem 4.3.

gives us the desired results along generic directions (non-normal lines of the polytope), but is not quite as powerful along critical directions (normal lines of the polytope), where it predicts that a typical graph is not unique but rather chosen by some unspecified distribution from two complete multipartite graphs. We then resort to a second approach, the finite  $n$  analysis. We interchange the order of limits in this approach: the parameter  $r$  (which quantifies the magnitude of  $\beta_2$ ) goes to infinity first followed by the growth of the system size  $n$ . Even though the asymptotics considered in these two approaches are not exactly the same, the findings are almost identical, except that the first approach leaves some ambiguity over the smoothness of the transition along critical directions due to the unspecified limiting probability distribution (as per Theorem 3.2 in [10]), while the second approach provides a deterministic result along critical directions and shows that the limiting behavior of the model depends on the natural parametrization in a discontinuous manner.

We have adopted different notations for these two approaches. In the variational analysis, we take the straight line as  $\beta_1 = a\beta_2 + b$ , where  $a$  and  $b$  are fixed constants (corresponding to slope and intercept of the line, respectively) and  $\beta_2$  decays to negative infinity. In the finite  $n$  analysis, we represent the straight line in vector format  $\beta + ro$ , where  $\beta$  and  $o$  are fixed vectors in  $\mathbb{R}^2$  and  $r$  diverges to infinity. The slope-intercept form is well suited for the infinite dimensional calculation (see Section 3), and the vector format is more natural for the finite dimensional calculation (see Section 4), where the geometry plays a bigger role in the proofs, so we keep them separate. Nevertheless, we would like to point out that there is a natural correspondence between these two formulations: the slope  $a$  corresponds to the direction  $o$  and the intercept  $b$  corresponds to the initial value  $\beta$ . For both infinite and finite dimensional calculations, the structure of the graph is mainly determined by  $a$ . Indeed, when  $a$  is associated with generic directions (non-normal lines), or equivalently,  $a_k > a > a_{k+1}$  (see Theorem 3.3),  $b$  does not matter (see Theorem 4.2). But when  $a$  is associated with critical directions  $o_k$  (normal lines), or equivalently,  $a = a_k$  (see Theorem 3.3),  $b$  plays a critical role.  $b \geq 0$  corresponds to  $\beta$  in  $H_k^+$  or  $H_k$ , and so a typical graph is  $(k+2)$ -partite, while  $b < 0$  corresponds to  $\beta$  in  $H_k^-$ , and so a typical graph is  $(k+1)$ -partite (see Theorem 4.3). We see that finite dimensional

analysis (though only completely rigorous for certain  $n$ 's) provides supporting evidence for the variational analysis.

Our focus is on the “repulsive” 2-parameter exponential random graph model, but the quantized behavior discovered in the repulsive region may be extended to the attractive region also. In [46], it was shown that the “attractive” 2-parameter exponential random graph model contains a continuous curve across which the graph exhibits jump discontinuity in density. For large and positive values of  $\beta_2$ , this curve asymptotically approaches the straight line  $\beta_1 = -\beta_2$ . As  $\beta_2$  varies from below the curve to above the curve, the model transitions from being a very sparse graph to a very dense (nearly complete) graph, completely skipping all intermediate structures. Theorems 3.1 and 4.4 capture this intriguing behavior.

Lastly, we would like to state that we do not actually need the exact expressions of the lower boundary of homomorphism densities (the Razborov curve) to derive our results, all we need is its strict concavity. According to Bollobás [4, 5], the vertices of the convex hull  $P$  for  $K_2$  and  $K_n$ , not just  $K_2$  and  $K_3$  as in the edge-triangle case, are given by the limits of complete  $k$ -equipartite graphs. More general conjectures of the limiting object  $P$  may be found in [18]. Similar asymptotic quantization phenomenon is thus expected for more generic “repulsive” models.

#### ACKNOWLEDGEMENTS

The authors benefited from participation in the 2013 workshop on Exponential Random Network Models at the American Institute of Mathematics. They thank the anonymous referees for their useful comments and suggestions and Ed Scheinerman of the Johns Hopkins University for making available the Matgraph toolbox that was used as a basis for many of their MATLAB simulations. Mei Yin's research was partially supported by NSF grant DMS-1308333. She gratefully acknowledges useful discussions with Charles Radin and Lorenzo Sadun. Alessandro Rinaldo's research was partially supported by grant FA9550-12-1-0392 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA) and by NSF CAREER grant DMS-1149677. He is grateful to Kayvan Sadeghi for discussions and help with early drafts of the manuscript.

#### 6. PROOFS

*Proof of Theorem 3.1.* Suppose  $H_2$  has  $p$  edges. Subject to  $\beta_1 = a\beta_2 + b$ , the variational problem (2.6) in the Erdős-Rényi region takes the following form: Find  $u$  so that

$$\beta_2(au + u^p) + bu - I(u) \tag{6.1}$$

is maximized. Take an arbitrary sequence  $\beta_2^{(i)} \rightarrow \infty$ . Let  $u_i$  be a maximizer corresponding to  $\beta_2^{(i)}$  and  $u^*$  be a limit point of the sequence  $\{u_i\}$ . By the boundedness of  $bu$  and  $I(u)$ , we see that  $u^*$  must maximize  $au + u^p$ . For  $a \neq -1$ , this maximizer is unique, but for  $a = -1$ , both 0 and 1 are maximizers. In this case,  $\beta_2(au + u^p) = 0$ , so we check the value of  $bu - I(u)$  as well. We conclude that  $u^* = 1$  for  $b > 0$ ,  $u^* = 0$  for  $b < 0$ , and  $u^*$  may be either 1 or 0 for  $b = 0$ .  $\square$

The following lemma appeared as an exercise in [39].

**Lemma 6.1** (Lovász). *Let  $F$  and  $G$  be two simple graphs. Let  $f$  be a graphon such that  $t(F, G) > 0$  and  $t(G, f) > 0$ . Then  $t(F, f) > 0$ .*

*Proof.* The proof of this lemma is an application of the Lebesgue density theorem. Suppose  $|V(F)| = m$  and  $|V(G)| = n$ . Since  $t(G, f) > 0$ , there is a Lebesgue measurable set  $A \subseteq \mathbb{R}^n$

and  $|A| \neq 0$  such that  $\prod_{\{i,j\} \in E(G)} f(x_i, x_j) > 0$  for  $x \in A$ . Since  $t(F, G) > 0$ , there exists a graph homomorphism  $h : V(F) \rightarrow V(G)$ . The preimage of  $A$  under  $h$  is then a Lebesgue measurable set  $B \subseteq \mathbb{R}^m$  and  $|B| \neq 0$  such that  $\prod_{\{i,j\} \in E(F)} f(y_i, y_j) > 0$  for  $y \in B$ . It follows that  $t(F, f) > 0$ .  $\square$

*Proof of Theorem 3.2.* Take an arbitrary sequence  $\beta_2^{(i)} \rightarrow -\infty$ . For each  $\beta_2^{(i)}$ , we examine the corresponding variational problem (2.3). Let  $\tilde{f}_i$  be an element of  $\tilde{F}^*(\beta_2^{(i)})$ . Let  $\tilde{f}^*$  be a limit point of  $\tilde{f}_i$  in  $\tilde{\mathcal{W}}$  (its existence is guaranteed by the compactness of  $\mathcal{W}$ ). Suppose  $t(H_2, \tilde{f}^*) > 0$ . Then by the continuity of  $t(H_2, \cdot)$  and the boundedness of  $t(H_1, \cdot)$  and  $\iint_{[0,1]^2} I(\cdot) dx dy$ ,  $\lim_{i \rightarrow \infty} \psi_\infty(\beta_2^{(i)}) = -\infty$ . But this is impossible since  $\psi_\infty(\beta_2^{(i)})$  is uniformly bounded below, as can be easily seen by considering (3.2) as a test function. Thus  $t(H_2, \tilde{f}^*) = 0$ . Denote by  $K_r$  a complete graph on  $r$  vertices. Since  $H_2$  has chromatic number  $r$ ,  $t(H_2, K_r) > 0$ , which implies that  $t(K_r, \tilde{f}^*) = 0$  by Lemma 6.1. By Turán's theorem for  $K_r$ -free graph, the edge density  $e$  of  $\tilde{f}^*$  must satisfy  $e = t(H_1, \tilde{f}^*) \leq (r-2)/(r-1)$ . This implies that the measure of the set  $\{(x, y) \in [0, 1]^2 | \tilde{f}^*(x, y) > 0\}$  is at most  $(r-2)/(r-1)$ . Otherwise, the graphon  $\tilde{f}(x, y) = \begin{cases} 1 & \tilde{f}^*(x, y) > 0; \\ 0 & \text{otherwise} \end{cases}$  would be  $K_r$ -free but with edge density greater than  $(r-2)/(r-1)$ , which is impossible.

Take an arbitrary edge density  $e \leq (r-2)/(r-1)$ . We consider all graphons  $f$  such that  $t(H_1, f) = e$  and  $t(H_2, f) = 0$ . Subject to these constraints, maximizing (2.3) is equivalent to minimizing  $\iint_{[0,1]^2} I(f) dx dy$ . Since  $t(H_2, f) = 0$ , as argued above, the set  $A = \{(x, y) \in [0, 1]^2 | f(x, y) > 0\}$  has measure at most  $(r-2)/(r-1)$ . If the measure of  $A$  is less than  $(r-2)/(r-1)$ , we randomly group part of the set  $[0, 1]^2 - A$  into  $A$  so that the measure of  $A$  is exactly  $(r-2)/(r-1)$ . We note that

$$\iint_{[0,1]^2} I(f(x, y)) dx dy = \iint_A I(f(x, y)) dx dy. \quad (6.2)$$

More importantly, since  $I(\cdot)$  is convex, by Jensen's inequality, we have

$$\iint_A I(f(x, y)) dx dy \geq \frac{r-2}{r-1} I\left(\iint_A \frac{r-1}{r-2} f(x, y) dx dy\right) = \frac{r-2}{r-1} I\left(\frac{r-1}{r-2} e\right), \quad (6.3)$$

where the first equality is obtained only when  $f(x, y) \equiv e(r-1)/(r-2)$  on  $A$ .

The variational problem (2.3) is now further reduced to the following: Find  $e \leq (r-2)/(r-1)$  (and hence  $f(x, y)$ ) so that

$$\beta_1 e - \frac{r-2}{r-1} I\left(\frac{r-1}{r-2} e\right) \quad (6.4)$$

is maximized. Simple computation yields  $e = p(r-2)/(r-1)$ , where  $p = e^{2\beta_1}/(1 + e^{2\beta_1})$ . Thus  $pg$  (where  $g$  is defined in (3.2)) is a maximizer for (2.3) as  $\beta_2 \rightarrow -\infty$ . We claim that any other maximizer  $h$  (if it exists) must lie in the same equivalence class. Recall that  $h$  must be  $K_r$ -free. Also,  $h$  is zero on a set of measure  $1/(r-1)$  and is  $p$  on a set of measure  $(r-2)/(r-1)$ . The graphon  $\bar{h}(x, y) = \begin{cases} 1 & h(x, y) = p; \\ 0 & \text{otherwise} \end{cases}$  describes a  $K_r$ -free graph with edge density  $(r-2)/(r-1)$ . By Turán's theorem,  $\bar{h}$  corresponds to the complete  $(r-1)$ -equipartite graph, and is thus equivalent to  $g$ . Hence  $h = p\bar{h}$  is equivalent to  $pg$ .  $\square$

*Proof of Theorem 3.3.* Subject to  $\beta_1 = a\beta_2 + b$ , the variational problem (2.3) takes the following form: Find  $f(x, y)$  so that

$$\beta_2(ae + t) + be - \iint_{[0,1]^2} I(f(x, y)) dx dy \quad (6.5)$$

is maximized, where  $e = t(H_1, f)$  denotes the edge density and  $t = t(H_2, f)$  denotes the triangle density of  $f$ , respectively. Take an arbitrary sequence  $\beta_2^{(i)} \rightarrow -\infty$ . Let  $\tilde{f}_i$  be an element of  $\tilde{F}^*(\beta_2^{(i)})$ . Let  $\tilde{f}^*$  be a limit point of  $\tilde{f}_i$  in  $\tilde{\mathcal{W}}$  (its existence is guaranteed by the compactness of  $\mathcal{W}$ ). By the continuity of  $t(H_2, \cdot)$  and the boundedness of  $t(H_1, \cdot)$  and  $\iint_{[0,1]^2} I(\cdot) dx dy$ , we see that  $f^*$  must minimize  $ae + t$ . This implies that  $f^*$  must lie on the Razborov curve (i.e., lower boundary of the feasible region) (see Figure 1). For every possible minimizer  $f$ ,  $t$  is thus an explicit function of  $e$ , given by the right-hand side of (2.14). The derivative of  $ae + t$  with respect to  $e$  is given by

$$a + \frac{3(k-1)}{k(k+1)} \left( k + \sqrt{k(k - e(k+1))} \right). \quad (6.6)$$

It is a decreasing function of  $e$  on each subinterval  $[(k-1)/k, k/(k+1)]$ , hence we further conclude that the minimizer  $f$  can only be obtained at the connection points  $e_k = k/(k+1)$ .

Consider two adjacent connection points  $(e_k, t_k)$  and  $(e_{k+1}, t_{k+1})$ , where

$$(e_k, t_k) = \left( \frac{k}{k+1}, \frac{k(k-1)}{(k+1)^2} \right) \text{ and } (e_{k+1}, t_{k+1}) = \left( \frac{k+1}{k+2}, \frac{k(k+1)}{(k+2)^2} \right). \quad (6.7)$$

Let  $L_k$  be the line segment joining these two points. The slope of the line passing through  $L_k$  is

$$\frac{k(3k+5)}{(k+1)(k+2)} = -a_k. \quad (6.8)$$

It is clear that  $a_k$  is a decreasing function of  $k$  and  $a_k \rightarrow -3$  as  $k \rightarrow \infty$ . More importantly, if  $a > a_k$ , then  $ae_k + t_k < ae_{k+1} + t_{k+1}$ ; if  $a = a_k$ , then  $ae_k + t_k = ae_{k+1} + t_{k+1}$ ; and if  $a < a_k$ , then  $ae_k + t_k > ae_{k+1} + t_{k+1}$ . Decreasing  $a$  thus moves the location of the minimizer  $f$  upward along the Razborov curve, with sudden jumps happening at special angles  $a = a_k$ , where the sign of  $b$  comes into play as in the proof of Theorem 3.1.  $\square$

*Proof of Lemma 4.1.* For part 1., the proof of Theorem 2 in [4] implies that any linear functional of the form  $L_\gamma(x) = \langle x, c \rangle$ , where  $c = (1, \gamma)^\top$  with  $\gamma \in \mathbb{R}$ , is maximized over  $P_n$  by some  $v_{k,n}$  and, conversely, any point  $v_{k,n}$  is such that

$$v_{k,n} = \operatorname{argmax}_{x \in P_n} L_\gamma(x) \quad (6.9)$$

for some  $\gamma \in \mathbb{R}$ . Thus,  $P_n$  is the convex hull of the points  $\{v_{k,n}, k = 0, 1, \dots, n-1\}$ . Next, if  $r \geq \lceil n/2 \rceil$ , the size of the larger class(es) of any  $T(n, r)$  is 2 and the size of the smaller class(es) (if any) is 1. Thus, the increase in the number of edges and triangles going from  $T(n, r)$  to  $T(n, r+1)$  is 1 and  $(n-2)$ , respectively. As a result, the points  $\{t(T(n, r)), r = \lceil n/2 \rceil, \dots, n\}$  are collinear.

To show part 2., notice that, by definition,  $P_n \subset P$ , so it is enough to show that for any  $x \in P$  and  $\epsilon > 0$  there exists an  $n' = n'(x, \epsilon)$  such that  $\inf_{y \in P_n} \|x - y\| < \epsilon$  for all  $n > n'$ . But this follows from the fact that, for each fixed  $k$ ,  $\lim_{n \rightarrow \infty} v_{k,n} = v_k$  and every  $x \in P$  is either an extreme point of  $P$  or is contained in the convex hull of a *finite* number of extreme points of  $P$ .

The first claim of part 3. can be directly verified with easy algebra (see (2.8)). The second claim follows from Theorem 4.1 in [47] and the strict concavity of the lower boundary of  $R$  on each subinterval  $[(k-1)/k, k/(k+1)]$ .  $\square$

The key steps of the proofs of Theorems 4.2 and, in particular, 4.3 rely on a careful analysis of the closure of the exponential family corresponding to the model under study. For the sake of clarity, we will provide a self-contained treatment. For details, see [13, 14, 25, 48].

*The closure of  $\mathcal{E}_n$ .* Fix a positive integer  $k$ . We first describe the total variation closure of the family  $\mathcal{E}_n$  for  $n$  tending to infinity as  $n = j(k+1)(k+2)$ , for  $j = 1, 2, \dots$ . As a result, for all such  $n$ ,  $v_{k,n} = v_k$  and  $v_{k+1,n} = v_{k+1}$ , which implies  $L_{k,n} = L_k$  (see Lemma 4.1).

Let  $\nu_{k,n}$  be the restriction of  $\nu_n$  to  $L_k$  and consider the exponential family on  $P_n \cap L_{k,n} = \{v_k, v_{k+1}\}$  generated by  $\nu_{k,n}$  and  $t$ , and parametrized by  $\mathbb{R}^2$ , which we denote with  $\mathcal{E}_{k,n}$ . Thus, the probability of observing the point  $x \in \{v_k, v_{k+1}\}$  is

$$\mathbb{P}_{n,k,\beta}(x) = \frac{e^{n^2\langle x, \beta \rangle}}{e^{n^2\langle v_k, \beta \rangle} \nu_n(v_k) + e^{n^2\langle v_{k+1}, \beta \rangle} \nu_n(v_{k+1})} \nu_n(x), \quad \beta \in \mathbb{R}^2. \quad (6.10)$$

The new family  $\mathcal{E}_{k,n}$  is an element of the closure of  $\mathcal{E}_n$  in the topology corresponding to the variation metric. More precisely, the family  $\mathcal{E}_{k,n}$  is comprised by all the limits in total variation of sequence of distributions from  $\mathcal{E}_n$  parametrized by sequences  $\{\beta^{(i)}\} \subset \mathbb{R}^2$  such that  $\lim_i \|\beta^{(i)}\| = \infty$  and  $\lim_i \frac{\beta^{(i)}}{\|\beta^{(i)}\|} \rightarrow \frac{o_k}{\|o_k\|}$ .

**Proposition 6.2.** *Let  $n$  be fixed and a multiple of  $(k+1)(k+2)$ . For any  $\beta \in \mathbb{R}^2$ , consider the sequence of parameters  $\{\beta^{(i)}\}_{i=1,2,\dots}$  given by  $\beta^{(i)} = \beta + r_i o_k$ , where  $\{r_i\}_{i=1,2,\dots}$  is a sequence of positive numbers tending to infinity. Then,*

$$\lim_i \mathbb{P}_{n,\beta^{(i)}}(x) = \begin{cases} \mathbb{P}_{n,k,\beta}(x) & \text{if } x \in \{v_k, v_{k+1}\} \\ 0 & \text{if } x \in S_n \setminus \{v_k, v_{k+1}\} \end{cases}$$

In particular,  $\mathbb{P}_{n,\beta^{(i)}}$  converges in total variation to  $\mathbb{P}_{n,k,\beta}$  as  $i \rightarrow \infty$ .

*Proof.* The proof can be found in, e.g., [13, 48]. We provide it for completeness.

Let  $x^* \in S_n$ . Then, for any  $\beta \in \mathbb{R}^2$ ,

$$\lim_{i \rightarrow \infty} \mathbb{P}_{n,\beta^{(i)}}(x^*) = \frac{e^{n^2\langle x^*, \beta \rangle}}{\lim_{i \rightarrow \infty} e^{n^2\psi_n(\beta^{(i)}) - r_i n^2\langle x^*, o_k \rangle}} \nu_n(x^*).$$

First suppose that  $x^* \in \{v_k, v_{k+1}\}$ . Then ,

$$\begin{aligned} & e^{n^2\psi_n(\beta^{(i)}) - r_i n^2\langle x^*, o_k \rangle} \\ &= \sum_{x \in S_n \setminus \{v_k, v_{k+1}\}} e^{n^2\langle x, \beta \rangle + r_i n^2\langle x - x^*, o_k \rangle} \nu_n(x) + \sum_{x \in \{v_k, v_{k+1}\}} e^{n^2\langle x, \beta \rangle} \nu_n(x) \\ & \quad \downarrow e^{n^2\langle v_k, \beta \rangle} \nu_n(v_k) + e^{n^2\langle v_{k+1}, \beta \rangle} \nu_n(v_{k+1}), \end{aligned}$$

as  $i \rightarrow \infty$ , because  $\sum_{x \in S_n \setminus \{v_k, v_{k+1}\}} e^{n^2\langle x, \beta \rangle + r_i n^2\langle x - x^*, o_k \rangle} \nu_n(x) \downarrow 0$ . This follows easily from the dominated convergence theorem, since the term  $\langle x - x^*, o_k \rangle$  is 0 if  $x \in \{v_k, v_{k+1}\}$  and is strictly negative otherwise. Thus,  $\mathbb{P}_{n,\beta^{(i)}}(x^*)$  converges to  $\mathbb{P}_{n,k,\beta}(x^*)$  (see (6.10)).

If  $x^* \in S_n \setminus \{v_k, v_{k+1}\}$ , since  $P_n$  is full dimensional, we have instead

$$e^{n^2\psi_n(\beta^{(i)}) - r_i n^2\langle x^*, o_k \rangle} \geq \sum_{x \in S_n : \langle x - x^*, o_k \rangle > 0} e^{n^2\langle x, \beta \rangle + r_i n^2\langle x - x^*, o_k \rangle} \nu_n(x) \rightarrow \infty,$$

as  $i \rightarrow \infty$ , by the monotone convergence theorem. Therefore  $\mathbb{P}_{n,\beta^{(i)}}(x^*) \rightarrow 0$ . The proof is now complete.  $\square$



The parametrization (6.10) is thus redundant, as it requires two parameters to represent a distribution whose support lies on a 1-dimensional hyperplane. One parameter is all that is needed to describe this distribution, a reduction that can be accomplished by standard arguments. Because such reparametrization is highly relevant to our problem, we provide the details.

**Proposition 6.3.** *The family  $\mathcal{E}_{k,n}$  is a one-dimensional exponential family parametrized by  $\mathcal{L}_k$ . Equivalently,  $\mathcal{E}_{k,n}$  can be parametrized with  $\{\langle l_k, \beta \rangle, \beta \in \mathbb{R}^2\} = \mathbb{R}$  as follows:*

$$\mathbb{P}_{n,k,\beta}(x) = \frac{e^{n^2 \langle x, l_k \rangle \cdot \langle l_k, \beta \rangle}}{e^{n^2 \langle v_k, l_k \rangle \cdot \langle l_k, \beta \rangle} \nu_n(v_k) + e^{n^2 \langle v_{k+1}, l_k \rangle \cdot \langle l_k, \beta \rangle} \nu_n(v_{k+1})} \nu_n(x), \quad (6.11)$$

where  $x \in \{v_k, v_{k+1}\}$ .

*Proof.* For an  $x \in \mathbb{R}^2$  and a linear subspace  $\mathcal{S}$  of  $\mathbb{R}^2$ , let  $\Pi_{\mathcal{S}}(x)$  be the orthogonal projection of  $x$  onto  $\mathcal{S}$  with respect to the Euclidean metric. Set  $\tilde{o}_k = \frac{o_k}{\|o_k\|}$  and let  $\alpha_k \in \mathbb{R}$  define the one-dimensional hyperplane (i.e., the line) going through  $L_k$ , i.e.,  $\{x \in \mathbb{R}^2: \langle x, \tilde{o}_k \rangle = \alpha_k\}$ . Then for every  $\beta \in \mathbb{R}^2$  and  $x \in \{v_k, v_{k+1}\}$ , we have

$$\langle x, \beta \rangle = \langle \Pi_{\mathcal{L}_k} x, \beta \rangle + \langle \Pi_{\mathcal{L}_k^\perp} x, \beta \rangle = \langle x, l_k \rangle \cdot \langle l_k, \beta \rangle + \alpha_k \langle \tilde{o}_k, \beta \rangle$$

since  $\alpha_k = \langle v_k, \tilde{o}_k \rangle = \langle v_{k+1}, \tilde{o}_k \rangle$ . Plugging into (6.10), we obtain (6.11). From that equation we see that, for any pair of distinct parameter vectors  $\beta$  and  $\beta'$ ,  $\mathbb{P}_{n,k,\beta} = \mathbb{P}_{n,k,\beta'}$  if and only if  $\langle l_k, \beta \rangle = \langle l_k, \beta' \rangle$ , i.e., if and only if they project to the same point in  $\mathcal{L}_k$ . This proves the claim.  $\square$

**Remark.** *The geometric interpretation of Proposition 6.3 is the following:  $\beta$  and  $\beta'$  parametrize the same distribution on  $\mathcal{E}_{k,n}$  if and only if the line going through them is parallel to the line spanned by  $o_k$ .*

Finally, the same arguments used in the proof of Proposition 6.2 also imply that the closure of  $\mathcal{E}_n$  along generic (i.e., non-critical) directions is comprised of point masses at the points  $v_{k,n}$ . For the next result, we do not need the condition of  $n$  being multiple of  $(k+1)(k+2)$ .

**Corollary 6.4.** *Let  $o \in \mathbb{R}^2$  be different from  $o_j$ ,  $j = -1, 0, 1, \dots$  and let  $k$  be such that  $o \in C_k^\circ$ . There exists an  $n_0 = n_0(o)$  such that, for any fixed  $n > n_0$ , and any sequence of parameters  $\{\beta^{(i)}\}_{i=1,2,\dots}$  given by  $\beta^{(i)} = \beta + r_i o$ , where  $\{r_i\}_{i=1,2,\dots}$  is a sequence of positive numbers tending to infinity and  $\beta$  is a vector in  $\mathbb{R}^2$ ,*

$$\lim_i \mathbb{P}_{n,\beta^{(i)}}(x) = \begin{cases} 1 & \text{if } x = v_{k,n} \\ 0 & \text{otherwise.} \end{cases}$$

*That is,  $\mathbb{P}_{n,\beta^{(i)}}$  converges in total variation to the point mass at  $v_k$  as  $i \rightarrow \infty$ .*

*Proof.* We only provide a brief sketch of the proof. From Lemma 4.1,  $P_n$  is the convex hull of the points  $\{v_{k,n}, k = 0, 1, \dots, \lceil n/2 \rceil - 1\}$  and  $v_{n-1,n}$  and, for each fixed  $k$ ,  $v_{k,n} \rightarrow v_k$  as  $n \rightarrow \infty$ . Therefore, the normal cone to  $v_{k,n}$  converges to  $C_k$ . Since by assumption  $o \in C_k^\circ$ , there exists an  $n_0$ , which depends on  $o$  (and hence also on  $k$ ), such that, for all  $n > n_0$ ,  $o$  is in the interior of the normal cone to  $v_{k,n}$ . The arguments used in the proof of Proposition 6.2 yield the desired claim.  $\square$

*Asymptotics of the closure of  $\mathcal{E}_n$ .* We now study the asymptotic properties of the families  $\mathcal{E}_{k,n}$  for fixed  $k$  and as  $n$  tends to infinity as  $n = j(k+1)(k+2)$  for  $j = 1, 2, \dots$

**Theorem 6.5.** *Let  $\{n_j\}_{j=1,2,\dots}$  be the sequence  $n_j = j(k+1)(k+2)$ . Then,*

$$\lim_{j \rightarrow \infty} \frac{\mathbb{P}_{n_j, k, \beta}(v_{k+1})}{\mathbb{P}_{n_j, k, \beta}(v_k)} \rightarrow \begin{cases} \infty & \text{if } \beta \in H_k^+ \text{ or } \beta \in H_k, \\ 0 & \text{if } \beta \in H_k^-. \end{cases}$$

**Remark.** *The proof further shows that the ratio of probabilities diverges or vanishes at a rate exponential in  $n_j^2$ .*

*Proof.* We can write

$$\frac{\mathbb{P}_{n_j, k, \beta}(v_{k+1})}{\mathbb{P}_{n_j, k, \beta}(v_k)} = e^{n^2 \langle l_k, \beta \rangle \langle v_{k+1} - v_k, l_k \rangle} \frac{\nu_n(v_{k+1})}{\nu_n(v_k)}.$$

We will first analyze the limiting behavior of the dominating measure  $\nu_n$ . We will show that, as  $n \rightarrow \infty$ , the number of Turán graphs with  $r+1$  classes is larger than the number of Turán graphs with  $r$  classes by a multiplicative factor that is exponential in  $n$ .

**Lemma 6.6.** *Consider the sequence of integers  $n = j(k+1)(k+2)$ , where  $k \geq 1$  is a fixed integer and  $j = 1, 2, \dots$ . Then, as  $n \rightarrow \infty$ ,*

$$\frac{\nu_n(v_{k+1})}{\nu_n(v_k)} \asymp \sqrt{\frac{1}{n}} \left( \frac{k+2}{k+1} \right)^n,$$

*Proof.* Recall that  $\nu_n(v_k)$  is the number of (simple, labeled) graphs on  $n$  nodes isomorphic to a Turán graph with  $(k+1)$  classes each of size  $j(k+2)$ , and that  $\nu_n(v_{k+1})$  is the number of (simple, labeled) graphs on  $n$  nodes isomorphic to a Turán graph with  $(k+2)$  classes each of size  $j(k+1)$ . Thus,

$$\begin{aligned} \nu_n(v_k) &= \frac{1}{(k+1)!} \frac{n!}{[(j(k+2))]^{k+1}}. \\ \nu_n(v_{k+1}) &= \frac{1}{(k+2)!} \frac{n!}{[(j(k+1))]^{k+2}}. \end{aligned}$$

Next, since  $n = j(k+1)(k+2)$ , using Stirling's approximation,

$$\begin{aligned} ((j(k+2)))^{k+1} &\sim (2\pi j(k+2))^{(k+1)/2} e^{-j(k+2)(k+1)} (j(k+2))^{j(k+2)(k+1)} \\ &= (2\pi j(k+2))^{(k+1)/2} e^{-n} (j(k+2))^n, \\ ((j(k+1)))^{k+2} &\sim (2\pi j(k+1))^{(k+2)/2} e^{-j(k+1)(k+2)} (j(k+1))^{j(k+1)(k+2)} \\ &= (2\pi j(k+1))^{(k+2)/2} e^{-n} (j(k+1))^n. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\nu_n(v_{k+1})}{\nu_n(v_k)} &\sim \frac{(k+1)! (2\pi)^{(k+1)/2} (j(k+2))^{(k+1)/2} (j(k+2))^n}{(k+2)! (2\pi)^{(k+2)/2} (j(k+1))^{(k+2)/2} (j(k+1))^n} \\ &= \left[ \frac{(k+1)! (2\pi)^{(k+1)/2}}{(k+2)! (2\pi)^{(k+2)/2}} \left( \frac{k+2}{k+1} \right)^{(k+1)/2} \sqrt{k+2} \right] \sqrt{\frac{1}{n}} \left( \frac{k+2}{k+1} \right)^n, \end{aligned}$$

where we have used the fact that  $j = \frac{n}{(k+1)(k+2)}$  for each  $n$ .  $\square$

Basic geometry considerations yield that, for any  $\beta \in \mathbb{R}^2$ ,

$$\langle l_k, \beta \rangle \begin{cases} > 0 & \text{if } \beta \in H_k^+ \\ < 0 & \text{if } \beta \in H_k^- \\ = 0 & \text{if } \beta \in H_k. \end{cases} \quad (6.12)$$

Next, we have that

$$\langle v_{k+1} - v_k, l_k \rangle > 0,$$

since

$$l_k = \frac{1}{\sqrt{1 + \left(\frac{k(3k+5)}{(k+1)(k+2)}\right)^2}} \begin{pmatrix} 1 \\ \frac{k(3k+5)}{(k+1)(k+2)} \end{pmatrix} \quad \text{and} \quad v_{k+1} - v_k = \begin{pmatrix} \frac{1}{(k+1)(k+2)} \\ \frac{k(3k+5)}{(k+1)^2(k+2)^2} \end{pmatrix}$$

are parallel vectors with positive entries.

By Lemma 6.6, we finally conclude that

$$\frac{\mathbb{P}_{n_j, k, \beta}(v_{k+1})}{\mathbb{P}_{n_j, k, \beta}(v_k)} \asymp e^{n^2 C_k(\beta)} \sqrt{\frac{1}{n}} \left(\frac{k+2}{k+1}\right)^n.$$

where  $C_k(\beta) = \langle l_k, \beta \rangle \langle v_{k+1} - v_k, l_k \rangle$ . The result now follows since the term  $e^{n^2 C_k(\beta)}$  dominates the other term and  $\text{sign}(C_k(\beta)) = \text{sign}(\langle l_k, \beta \rangle)$ .  $\square$

*Proofs of Theorems 4.2, 4.3, 4.4, and 4.5.* We first consider Theorem 4.3. Assume that  $\beta \in H_k^+$  or  $\beta \in H_k$ . Then by Theorem 6.5, there exists an  $n_0 = n_0(\beta, \epsilon, k)$  such that, for all  $n > n_0$  and a multiple of  $(k+1)(k+2)$ ,  $\mathbb{P}_{n, k, \beta}(v_{k+1}) > 1 - \epsilon/2$ . Let  $n$  be an integer larger than  $n_0$  and a multiple of  $(k+1)(k+2)$ . By Proposition 6.2, there exists an  $r_0 = r_0(\beta, \epsilon, k, n)$  such that, for all  $r > r_0$ ,  $\mathbb{P}_{n, \beta + r o_k}(v_{k+1}) > \mathbb{P}_{n, k, \beta}(v_{k+1}) - \epsilon/2$ . Thus, for these values of  $n$  and  $r$ ,

$$\mathbb{P}_{n, \beta + r o_k}(v_{k+1}) > \mathbb{P}_{n, k, \beta}(v_{k+1}) - \epsilon/2 > 1 - \epsilon/2 - \epsilon/2 = 1 - \epsilon,$$

as claimed. The case of  $\beta \in H_k^-$  is proved in the same way.

For Theorem 4.2, we use Lemma 6.4, which guarantees that there exists an integer  $n_0 = n_0(\beta, \epsilon, o)$  such that, for any integer  $n > n_0$ , there exists an  $r_0 = r_0(\beta, \epsilon, o, n)$  such that for any  $r > r_0$ ,

$$\mathbb{P}_{n, \beta + r o}(v_{k, n}) > 1 - \epsilon.$$

Theorem 4.4 is proved as a direct corollary of Proposition 6.3 along with simple algebra. Finally, Theorem 4.5 follows from similar arguments used in the proof of Proposition 6.3.  $\square$

## REFERENCES

- [1] Aldous, D.: Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* 11, 581-598 (1981)
- [2] Barndorff-Nielsen, O.: *Information and Exponential Families in Statistical Theory*. John Wiley & Sons, New York (1978)
- [3] Bhamidi, S., Bresler, G., Sly, A.: Mixing time of exponential random graphs. *Ann. Appl. Probab.* 21, 2146-2170 (2011)
- [4] Bollobás, B.: Relations between sets of complete subgraphs. In: Nash-Williams, C., Sheehan, J. (eds.) *Proceedings of the Fifth British Combinatorial Conference, Congressus Numerantium, No. XV*, pp. 79-84. Utilitas Mathematica Publishing, Winnipeg, Man (1976)
- [5] Bollobás, B.: *Random Graphs*, Volume 73 of Cambridge Studies in Advanced Mathematics. 2nd ed. Cambridge University Press, Cambridge (2001)

- [6] Borgs, C., Chayes, J., Lovász, L., Sós, V.T., Vesztergombi, K.: Counting graph homomorphisms. In: Klazar, M., Kratochvíl, J., Loeb, M., Thomas, R., Valtr, P. (eds.) *Topics in Discrete Mathematics, Volume 26*, pp. 315-371. Springer, Berlin (2006)
- [7] Borgs, C., Chayes, J.T., Lovász, L., Sós, V.T., Vesztergombi, K.: Convergent sequences of dense graphs I. Subgraph frequencies, metric properties and testing. *Adv. Math.* 219, 1801-1851 (2008)
- [8] Borgs, C., Chayes, J.T., Lovász, L., Sós, V.T., Vesztergombi, K.: Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math.* 176, 151-219 (2012)
- [9] Brown, L.: *Fundamentals of Statistical Exponential Families*. IMS Lecture Notes—Monograph Series, Volume 9. Institute of Mathematical Statistics, Hayward, CA (1986)
- [10] Chatterjee, S., Diaconis, P.: Estimating and understanding exponential random graph models. To appear in *Ann. Statist.* arXiv: 1102.2650 (2011)
- [11] Chatterjee, S., Diaconis, P., Sly, A.: Random graphs with a given degree sequence. *Ann. Appl. Probab.* 21, 1400-1435 (2011)
- [12] Chatterjee, S., Varadhan, S.R.S.: The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.* 32, 1000-1017 (2011)
- [13] Csiszár, I., Matúš, F.: Closure of exponential families. *Ann. Probab.* 33, 582-600 (2005)
- [14] Csiszár, I., Matúš, F.: Generalized maximum likelihood estimates for exponential families. *Probab. Theory Related Fields* 141, 213-246 (2008)
- [15] Diaconis, P., Janson, S.: Graph limits and exchangeable random graphs. *Rendiconti di Matematica* 28, 33-61 (2008)
- [16] Diestel, R.: *Graph Theory*. Springer, New York (2005)
- [17] Dorogovtsev, S., Mendes, J.: *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, Oxford (2003)
- [18] Engström, A., Norén, P.: Polytopes from subgraph statistics. arXiv: 1011.3552 (2011)
- [19] Erdős, P., Kleitman, D.J., Rothschild, B.L.: Asymptotic enumeration of  $K_n$ -free graphs. *International Colloquium on Combinatorial Theory, Volume 2*, pp. 19-27. *Atti dei Convegni Lincei*, Rome (1976)
- [20] Erdős, P., Rényi, A.: On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 5, 17-61 (1960)
- [21] Fienberg, S.: Introduction to papers on the modeling and analysis of network data. *Ann. Appl. Statist.* 4, 1-4 (2010)
- [22] Fienberg, S.: Introduction to papers on the modeling and analysis of network data II. *Ann. Appl. Statist.* 4, 533-534 (2010)
- [23] Fisher, D.C.: Lower bounds on the number of triangles in a graph. *J. Graph Theory* 13, 505-512 (1989)
- [24] Frank, O., Strauss, D.: Markov graphs. *J. Amer. Statist. Assoc.* 81, 832-842 (1986)
- [25] Geyer, C.J.: Likelihood inference in exponential families and directions of recession. *Electron. J. Stat.* 3, 259-289 (2009)
- [26] Geyer, C.J., Thompson, E.A.: Constrained Monte Carlo maximum likelihood for dependent data. *J. R. Statist. Soc. B* 54, 657-699 (1992)
- [27] Goldenberg, A., Zheng, A.X., Fienberg, S.E., Airoldi, E.M.: A survey of statistical network models. *Found. Trends Mach. Learn.* 2, 129-233 (2009)
- [28] Goodman, A.W.: On sets of acquaintances and strangers at any party. *Amer. Math. Monthly* 66, 778-783 (1959)
- [29] Haberman, S.: Discussion of an exponential family of probability distributions for directed graphs by Holland and Leinhardt. *J. Amer. Statist. Assoc.* 76, 60-61 (1981)
- [30] Häggström, O., Jonasson, J.: Phase transition in the random triangle model. *J. Appl. Probab.* 36, 1101-1115 (1999)
- [31] Handcock, M.S.: Assessing degeneracy in statistical models of social networks. Working paper no. 39, Center for Statistics and the Social Sciences, University of Washington (2003)
- [32] Holland, P., Leinhardt, S.: An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* 76, 33-50 (1981)
- [33] Hoover, D.: Row-column exchangeability and a generalized model for probability. In: Koch, G., Spizzichino, F. (eds.) *Exchangeability in Probability and Statistics*, pp. 281-291. North-Holland, Amsterdam (1982)
- [34] Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M., Morris, M.: *ergm*: a package to fit, simulate and diagnose exponential-family models for networks. *J. Statist. Softw.* 24, 1-29 (2008)
- [35] Kallenberg, O.: *Probabilistic Symmetries and Invariance Principles*. Springer, New York (2005)
- [36] Kolaczyk, E.D.: *Statistical Analysis of Network Data: Methods and Models*. Springer, New York (2009)

- [37] Lauritzen, S.L.: Rasch models with exchangeable rows and columns. In: Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M., West, M. (eds.) Bayesian Statistics 7, pp. 215-232. Oxford University Press, Oxford (2003)
- [38] Lovász, L.: Very large graphs. In: Jerison, D., Mazur, B., Mrowka, T., Schmid, W., Stanley, R., Yau, S.T. (eds.) Current Developments in Mathematics, Volume 2008, pp. 67-128. International Press, Boston (2009)
- [39] Lovász, L.: Large Networks and Graph Limits. American Mathematical Society, Providence (2012)
- [40] Lovász, L., Simonovits M.: On the number of complete subgraphs of a graph II. In: Erdős, P. et al. (eds.) Studies in Pure Mathematics, To the memory of Paul Turán, pp. 459-495. Springer, Basel (1983)
- [41] Lovász, L., Szegedy, B.: Limits of dense graph sequences. J. Combin. Theory Ser. B 96, 933-957 (2006)
- [42] Newman, M.: The structure and function of complex networks. SIAM Rev. 45, 167-256 (2003)
- [43] Newman, M.: Networks: An Introduction. Oxford University Press, New York (2010)
- [44] Pattison, P.E., Wasserman, S., 1999. Logit models and logistic regressions for social networks. II. Multivariate relations. British Journal of Mathematical and Statistical Psychology 52, 169-194.
- [45] Radin, C., Sadun, L.: Phase transitions in a complex network. J. Phys. A 46, 305002 (2013)
- [46] Radin, C., Yin, M.: Phase transitions in exponential random graphs. Ann. Appl. Probab. 23, 2458-2471 (2013)
- [47] Razborov, A.: On the minimal density of triangles in graphs. Combin. Probab. Comput. 17, 603-618 (2008)
- [48] Rinaldo, A., Fienberg, S., Zhou, Y.: On the geometry of discrete exponential families with application to exponential random graph models. Electron. J. Stat. 3, 446-484 (2009)
- [49] Rinaldo, A., Petrović, S., Fienberg, S.E.: Maximum likelihood estimation in network models. Ann. Statist. 41, 1085-1110 (2013)
- [50] Robins, G.L., Pattison, P.E., Kalish, Y., Lusher, D.: An introduction to exponential random graph ( $p^*$ ) models for social networks. Social Networks 29, 173-191 (2007)
- [51] Robins, G.L., Pattison, P.E., Wasserman, S.: Logit models and logistic regressions for social networks: III. Valued relations. Psychometrika 64, 371-394 (1999)
- [52] Shalizi, C.R., Rinaldo, A.: Consistency under sampling of exponential random graph models. Ann. Statist. 41, 508-535 (2013)
- [53] Snijders, T., Pattison, P., Robins, G., Handcock M.: New specifications for exponential random graph models. Sociol. Method. 36, 99-153 (2006)
- [54] Strogatz, S.: Exploring complex networks. Nature 410, 268-276 (2001)
- [55] Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (2010)
- [56] Wasserman, S., Pattison, P.E.: Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and  $p^*$ . Psychometrika 61, 401-425 (1996)
- [57] Watts, D., Strogatz, S.: Collective dynamics of 'small-world' networks. Nature 393, 440-442 (1998)
- [58] Yan, T., Xu, J.: A central limit theorem in the  $\beta$ -model for undirected random graphs with a diverging number of vertices. Biometrika 100, 519-524 (2013)

DEPARTMENT OF MATHEMATICS, BROWN UNIVERSITY, PROVIDENCE, RI 02912, USA  
*E-mail address:* [mei\\_yin@math.brown.edu](mailto:mei_yin@math.brown.edu)

DEPARTMENT OF STATISTICS, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PA 15213, USA  
*E-mail address:* [arinaldo@cmu.edu](mailto:arinaldo@cmu.edu)

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY, CAMBRIDGE, MA 02138, USA  
*E-mail address:* [sukhada23@gmail.com](mailto:sukhada23@gmail.com)