

# ON THE ASYMPTOTICS OF CONSTRAINED EXPONENTIAL RANDOM GRAPHS

RICHARD KENYON AND MEI YIN

June 17, 2014

ABSTRACT. The unconstrained exponential family of random graphs assumes no prior knowledge of the graph before sampling, but in many situations partial information of the graph is already known beforehand. A natural question to ask is what would be a typical random graph drawn from an exponential model subject to certain constraints? In particular, will there be a similar phase transition phenomenon as that which occurs in the unconstrained exponential model? We present some general results for the constrained model and then apply them to get concrete answers in the edge-triangle model.

## 1. INTRODUCTION

Consider the set  $\mathcal{G}_n$  of all simple graphs  $G_n$  on  $n$  vertices (“simple” means undirected, with no loops or multiple edges). By a  $k$ -parameter family of exponential random graphs we mean a family of probability measures  $\mathbb{P}_n^\beta$  on  $\mathcal{G}_n$  defined by, for  $G_n \in \mathcal{G}_n$ ,

$$\mathbb{P}_n^\beta(G_n) = \exp \left[ n^2 \left( \beta_1 t(H_1, G_n) + \cdots + \beta_k t(H_k, G_n) - \psi_n^\beta \right) \right], \quad (1.1)$$

where  $\beta = (\beta_1, \dots, \beta_k)$  are  $k$  real parameters,  $H_1, \dots, H_k$  are pre-chosen finite simple graphs (and we take  $H_1$  to be a single edge),  $t(H_i, G_n)$  is the density of graph homomorphisms (the probability that a random vertex map  $V(H_i) \rightarrow V(G_n)$  is edge-preserving),

$$t(H_i, G_n) = \frac{|\text{hom}(H_i, G_n)|}{|V(G_n)|^{|V(H_i)|}}, \quad (1.2)$$

and  $\psi_n^\beta$  is the normalization constant,

$$\psi_n^\beta = \frac{1}{n^2} \log \sum_{G_n \in \mathcal{G}_n} \exp \left[ n^2 (\beta_1 t(H_1, G_n) + \cdots + \beta_k t(H_k, G_n)) \right]. \quad (1.3)$$

Sometimes, other than homomorphism densities, we also consider more general bounded continuous functions on the graph space (a notion to be made precise later), for example the degree sequence or the eigenvalues of the adjacency matrix. These exponential models are commonly used to model real-world networks as they are able to capture a wide variety of common *network tendencies* by representing a complex global structure through a set of tractable local features [11, 15, 21]. Intuitively, we can think of the  $k$  parameters  $\beta_1, \dots, \beta_k$  as tuning parameters that allow one to adjust the influence of different subgraphs  $H_1, \dots, H_k$  of  $G_n$  on the probability distribution, whose asymptotics will be our main interest since networks are often very large in size.

---

Richard Kenyon’s research was partially supported by NSF grant DMS-1208191 and a Simons Investigator award. Mei Yin’s research was partially supported by NSF grant DMS-1308333.

Our main results are (Theorem 3.1) a variational principle for the normalization constant (partition function) for graphons with constrained edge density, and an associated concentration of measure (Theorem 3.2) indicating that almost all large constrained graphs lie near the maximizing set. We then specialize to the edge-triangle model, and show the existence of (first-order) phase transitions in the edge-density constrained models.

*Acknowledgements.* We thank Charles Radin, Kui Ren, and Lorenzo Sadun for helpful conversations.

## 2. BACKGROUND

We begin by reviewing some notation and results concerning the theory of graph limits and its use in exponential random graph models. Following the earlier work of Aldous [2] and Hoover [12], Lovász and coauthors (V.T. Sós, B. Szegedy, C. Borgs, J. Chayes, K. Vesztegombi,...) have constructed an elegant theory of graph limits in a sequence of papers [5, 6, 7, 14]. See also the recent book [13] for a comprehensive account and references. This sheds light on various topics such as graph testing and extremal graph theory, and has found applications in statistics and related areas (see for instance [9]). Though their theory has been developed for dense graphs (number of edges comparable to the square of number of vertices), serious attempts have been made at formulating parallel results for sparse graphs [3, 4].

Here are the basics of this beautiful theory. Any simple graph  $G_n$ , irrespective of the number of vertices, may be represented as an element  $h^{G_n}$  of a single abstract space  $\mathcal{W}$  that consists of all symmetric measurable functions from  $[0, 1]^2$  into  $[0, 1]$ , by defining

$$h^{G_n}(x, y) = \begin{cases} 1, & \text{if } (\lceil nx \rceil, \lceil ny \rceil) \text{ is an edge in } G_n; \\ 0, & \text{otherwise.} \end{cases} \quad (2.1)$$

A sequence of graphs  $\{G_n\}_{n \geq 1}$  is said to converge to a function  $h \in \mathcal{W}$  (referred to as a “graph limit” or “graphon”) if for every finite simple graph  $H$  with vertex set  $V(H) = [k] = \{1, \dots, k\}$  and edge set  $E(H)$ ,

$$\lim_{n \rightarrow \infty} t(H, h^{G_n}) = t(H, h), \quad (2.2)$$

where  $t(H, h^{G_n}) = t(H, G_n)$ , the graph homomorphism density (1.2), by construction, and

$$t(H, h) = \int_{[0,1]^k} \prod_{\{i,j\} \in E(H)} h(x_i, x_j) dx_1 \cdots dx_k. \quad (2.3)$$

Indeed every function in  $\mathcal{W}$  arises as the limit of a certain graph sequence [14]. Intuitively, the interval  $[0, 1]$  represents a “continuum” of vertices, and  $h(x, y)$  denotes the probability of putting an edge between  $x$  and  $y$ . For example, for the Erdős-Rényi random graph  $G(n, \rho)$ , the “graphon” is represented by the function that is identically equal to  $\rho$  on  $[0, 1]^2$ . This “graphon” interpretation enables us to capture the notion of convergence in terms of subgraph densities by an explicit metric on  $\mathcal{W}$ , the so-called “cut distance”:

$$d_{\square}(f, h) = \sup_{S, T \subseteq [0,1]} \left| \int_{S \times T} (f(x, y) - h(x, y)) dx dy \right| \quad (2.4)$$

for  $f, h \in \mathcal{W}$ . A non-trivial complication is that the topology induced by the cut metric is well defined only up to measure preserving transformations of  $[0, 1]$  (and up to sets of Lebesgue measure zero), which in the context of finite graphs may be thought of as vertex relabeling. To tackle this issue, an equivalence relation  $\sim$  is introduced in  $\mathcal{W}$ . We say that  $f \sim h$  if  $f(x, y) = h_{\sigma}(x, y) := h(\sigma x, \sigma y)$  for some measure preserving bijection  $\sigma$  of  $[0, 1]$ . Let  $\tilde{h}$  (referred

to as a “reduced graphon”) denote the equivalence class of  $h$  in  $(\mathcal{W}, d_\square)$ . Since  $d_\square$  is invariant under  $\sigma$ , one can then define on the resulting quotient space  $\tilde{\mathcal{W}}$  the natural distance  $\delta_\square$  by  $\delta_\square(\tilde{f}, \tilde{h}) = \inf_{\sigma_1, \sigma_2} d_\square(f_{\sigma_1}, h_{\sigma_2})$ , where the infimum ranges over all measure preserving bijections  $\sigma_1$  and  $\sigma_2$ , making  $(\tilde{\mathcal{W}}, \delta_\square)$  into a metric space. With some abuse of notation we also refer to  $\delta_\square$  as the “cut distance”.

The space  $(\tilde{\mathcal{W}}, \delta_\square)$  enjoys many important properties that are essential for the study of exponential random graph models. For example, it is a compact space and homomorphism densities  $t(H, \cdot)$  are continuous functions on it.

For the purpose of this paper, two theorems from Chatterjee and Diaconis [8] (both based on a large deviation result established in Chatterjee and Varadhan [10]) merit some special attention. Together they connect the occurrence of a phase transition in the exponential model with the solution of a certain maximization problem. Their results are formulated in terms of general exponential models where the terms in the exponent defining the probability measure may contain functions on the graph space other than homomorphism densities, as alluded to at the beginning of this paper. Let  $T : \tilde{\mathcal{W}} \rightarrow \mathbb{R}$  be a bounded continuous function. Let the probability measure  $\mathbb{P}_n$  and the normalization constant  $\psi_n$  be defined as in (1.1) and (1.3), that is,

$$\mathbb{P}_n(G_n) = \exp\left(n^2(T(\tilde{h}^{G_n}) - \psi_n)\right), \quad (2.5)$$

$$\psi_n = \frac{1}{n^2} \log \sum_{G_n \in \mathcal{G}_n} \exp\left(n^2 T(\tilde{h}^{G_n})\right). \quad (2.6)$$

The first theorem (Theorem 3.1 in [8]) states that the limiting normalization constant  $\psi := \lim_{n \rightarrow \infty} \psi_n$  of the exponential random graph, which is crucial for the computation of maximum likelihood estimates, always exists and is given by

$$\psi = \sup_{\tilde{h} \in \tilde{\mathcal{W}}} \left(T(\tilde{h}) - I(\tilde{h})\right), \quad (2.7)$$

where  $I$  is first defined as a function from  $[0, 1]$  to  $\mathbb{R}$  as

$$I(u) = \frac{1}{2}u \log u + \frac{1}{2}(1-u) \log(1-u), \quad (2.8)$$

and then extended to  $\tilde{\mathcal{W}}$  in the usual manner:

$$I(\tilde{h}) = \int_{[0,1]^2} I(h(x,y)) dx dy, \quad (2.9)$$

where  $h$  is any representative element of the equivalence class  $\tilde{h}$ . It was shown in [10] that  $I$  is well defined and lower semi-continuous on  $\tilde{\mathcal{W}}$ . Let  $\tilde{H}$  be the subset of  $\tilde{\mathcal{W}}$  where  $\psi$  is maximized. By the compactness of  $\tilde{\mathcal{W}}$ , the continuity of  $T$  and the lower semi-continuity of  $I$ ,  $\tilde{H}$  is a nonempty compact set. The set  $\tilde{H}$  encodes important information about the exponential model (2.5) and helps to predict the behavior of a typical random graph sampled from this model. The second theorem (Theorem 3.2 in [8]) states that in the large  $n$  limit, the quotient image  $\tilde{h}^{G_n}$  of a random graph  $G_n$  drawn from (2.5) must lie close to  $\tilde{H}$  with high probability,

$$\delta_\square(\tilde{h}^{G_n}, \tilde{H}) \rightarrow 0 \text{ in probability as } n \rightarrow \infty. \quad (2.10)$$

Since the limiting normalization constant  $\psi$  is the generating function for the limiting expectations of other random variables on the graph space such as expectations and correlations of homomorphism densities, a phase transition occurs when  $\psi$  is non-analytic or when  $\tilde{H}$  is not a singleton set.

## 3. CONSTRAINED EXPONENTIAL RANDOM GRAPHS

The exponential family of random graphs introduced above assumes no prior knowledge of the graph before sampling. But in many situations partial information of the graph is already known beforehand. For example, practitioners might be told that the edge density of the graph is close to  $1/2$  or the triangle density is close to  $1/4$  or the adjacency matrix of the graph obeys a certain form. A natural question to ask then is what would be a typical random graph drawn from an exponential model subject to these constraints? Or perhaps more importantly will there be a similar phase transition phenomenon as in the unconstrained exponential model? The following Theorems 3.1 and 3.2 give a detailed answer to these questions. Not surprisingly, the proofs follow a similar line of reasoning as in Theorems 3.1 and 3.2 of [8]. However, due to the imposed constraints, instead of working with probability measure  $\mathbb{P}_n$  and normalization constant  $\psi_n$  as in [8], we are working with conditional probability measure and conditional normalization constant, so the argument is more involved. The proof of Theorem 3.1 also incorporates some ideas from Theorem 3.1 of [17].

For clarity, we assume that the edge density of the graph is approximately known, though the proof runs through without much modification for more general constraints. We make precise the notion of “approximately known” below. We still assign a probability measure  $\mathbb{P}_n$  as in (2.5) on  $\mathcal{G}_n$ , but we will consider a conditional normalization constant and also define a conditional probability measure. Let  $e \in [0, 1]$  be a real parameter that signifies an “ideal” edge density. Take  $\alpha > 0$ . The conditional normalization constant  $\psi_{n,\alpha}^e$  is defined analogously to the normalization constant for the unconstrained exponential random graph model,

$$\psi_{n,\alpha}^e = \frac{1}{n^2} \log \sum_{G_n \in \mathcal{G}_n: |e(G_n) - e| < \alpha} \exp\left(n^2 T(\tilde{h}^{G_n})\right), \quad (3.1)$$

the difference being that we are only taking into account graphs  $G_n$  whose edge density  $e(G_n)$  is within an  $\alpha$  neighborhood of  $e$ . Correspondingly, the associated conditional probability measure  $\mathbb{P}_{n,\alpha}^e(G_n)$  is given by

$$\mathbb{P}_{n,\alpha}^e(G_n) = \exp(-n^2 \psi_{n,\alpha}^e) \exp\left(n^2 T(\tilde{h}^{G_n})\right) \mathbb{1}_{|e(G_n) - e| < \alpha}. \quad (3.2)$$

We perform two limit operations on  $\psi_{n,\alpha}^e$ . First we take  $n$  to infinity, then we shrink the interval around  $e$  by letting  $\alpha$  go to zero:

$$\psi^e = \lim_{\alpha \rightarrow 0} \lim_{n \rightarrow \infty} \psi_{n,\alpha}^e. \quad (3.3)$$

Intuitively, these two operations ensure that we are examining the asymptotics of exponentially weighted large graphs with edge density sufficiently close to  $e$ . Theorem 3.1 shows that this is indeed the case.

**Theorem 3.1.** *Let  $e : 0 \leq e \leq 1$  be a real parameter and  $T : \tilde{\mathcal{W}} \rightarrow \mathbb{R}$  be a bounded continuous function. Let  $I$  and  $\psi^e$  be defined as before (see (2.8), (2.9), (3.1) and (3.3)). Then*

$$\psi^e = \sup_{\tilde{h} \in \tilde{\mathcal{W}}: e(\tilde{h}) = e} \left( T(\tilde{h}) - I(\tilde{h}) \right), \quad (3.4)$$

where  $e(\tilde{h})$  is the edge density of the reduced graphon  $\tilde{h}$ , obtained by taking  $H$  to be a single edge in (2.3).

*Proof.* By definition  $\liminf \psi_{n,\alpha}^e$  and  $\limsup \psi_{n,\alpha}^e$  exist as  $n \rightarrow \infty$ . We will show that they both approach  $\sup_{\tilde{h}: e(\tilde{h}) = e} (T(\tilde{h}) - I(\tilde{h}))$  as  $\alpha \rightarrow 0$ . For this purpose we need to define a few sets. Let

$\tilde{U}_\alpha$  be the open strip of reduced graphons  $\tilde{h}$  with  $e - \alpha < e(\tilde{h}) < e + \alpha$ , and let  $\tilde{H}_\alpha$  be the closed strip  $e - \alpha \leq e(\tilde{h}) \leq e + \alpha$ . Fix  $\epsilon > 0$ . Since  $T$  is a bounded function, there is a finite set  $R$  such that the intervals  $\{(c, c + \epsilon) : c \in R\}$  cover the range of  $T$ . For each  $c \in R$ , let  $\tilde{U}_{\alpha,c}$  be the open set of reduced graphons  $\tilde{h}$  with  $e - \alpha < e(\tilde{h}) < e + \alpha$  and  $c < T(\tilde{h}) < c + \epsilon$ , and let  $\tilde{H}_{\alpha,c}$  be the closed set  $e - \alpha \leq e(\tilde{h}) \leq e + \alpha$  and  $c \leq T(\tilde{h}) \leq c + \epsilon$ . It may be assumed without loss of generality that  $\tilde{U}_{\alpha,c}$  and  $\tilde{H}_{\alpha,c}$  are nonempty for each  $c \in R$ . Let  $|\tilde{U}_{\alpha,c}^n|$  and  $|\tilde{H}_{\alpha,c}^n|$  denote the number of graphs with  $n$  vertices whose reduced graphons lie in  $\tilde{U}_{\alpha,c}$  or  $\tilde{H}_{\alpha,c}$ , respectively. The large deviation principle, Theorem 2.3 of [10], implies that:

$$\limsup_{n \rightarrow \infty} \frac{\log |\tilde{H}_{\alpha,c}^n|}{n^2} \leq - \inf_{\tilde{h} \in \tilde{H}_{\alpha,c}} I(\tilde{h}), \quad (3.5)$$

and that

$$\liminf_{n \rightarrow \infty} \frac{\log |\tilde{U}_{\alpha,c}^n|}{n^2} \geq - \inf_{\tilde{h} \in \tilde{U}_{\alpha,c}} I(\tilde{h}). \quad (3.6)$$

We first consider  $\limsup \psi_{n,\alpha}^e$ .

$$\exp(n^2 \psi_{n,\alpha}^e) \leq \sum_{c \in R} \exp(n^2(c + \epsilon)) |\tilde{H}_{\alpha,c}^n| \leq |R| \sup_{c \in R} \exp(n^2(c + \epsilon)) |\tilde{H}_{\alpha,c}^n|. \quad (3.7)$$

This shows that

$$\limsup_{n \rightarrow \infty} \psi_{n,\alpha}^e \leq \sup_{c \in R} \left( c + \epsilon - \inf_{\tilde{h} \in \tilde{H}_{\alpha,c}} I(\tilde{h}) \right). \quad (3.8)$$

Each  $\tilde{h} \in \tilde{H}_{\alpha,c}$  satisfies  $T(\tilde{h}) \geq c$ . Consequently,

$$\sup_{\tilde{h} \in \tilde{H}_{\alpha,c}} (T(\tilde{h}) - I(\tilde{h})) \geq \sup_{\tilde{h} \in \tilde{H}_{\alpha,c}} (c - I(\tilde{h})) = c - \inf_{\tilde{h} \in \tilde{H}_{\alpha,c}} I(\tilde{h}). \quad (3.9)$$

Substituting this in (3.8) gives

$$\begin{aligned} \limsup_{n \rightarrow \infty} \psi_{n,\alpha}^e &\leq \epsilon + \sup_{c \in R} \sup_{\tilde{h} \in \tilde{H}_{\alpha,c}} (T(\tilde{h}) - I(\tilde{h})) \\ &= \epsilon + \sup_{\tilde{h} \in \tilde{H}_\alpha} (T(\tilde{h}) - I(\tilde{h})). \end{aligned} \quad (3.10)$$

Next we consider  $\liminf \psi_{n,\alpha}^e$ .

$$\exp(n^2 \psi_{n,\alpha}^e) \geq \sup_{c \in R} \exp(n^2 c) |\tilde{U}_{\alpha,c}^n|. \quad (3.11)$$

Therefore for each  $c \in R$ ,

$$\liminf_{n \rightarrow \infty} \psi_{n,\alpha}^e \geq c - \inf_{\tilde{h} \in \tilde{U}_{\alpha,c}} I(\tilde{h}). \quad (3.12)$$

Each  $\tilde{h} \in \tilde{U}_{\alpha,c}$  satisfies  $T(\tilde{h}) < c + \epsilon$ . Therefore

$$\sup_{\tilde{h} \in \tilde{U}_{\alpha,c}} (T(\tilde{h}) - I(\tilde{h})) \leq \sup_{\tilde{h} \in \tilde{U}_{\alpha,c}} (c + \epsilon - I(\tilde{h})) = c + \epsilon - \inf_{\tilde{h} \in \tilde{U}_{\alpha,c}} I(\tilde{h}). \quad (3.13)$$

Together with (3.12), this shows that

$$\begin{aligned} \liminf_{n \rightarrow \infty} \psi_{n,\alpha}^e &\geq -\epsilon + \sup_{c \in R} \sup_{\tilde{h} \in \tilde{U}_{\alpha,c}} (T(\tilde{h}) - I(\tilde{h})) \\ &= -\epsilon + \sup_{\tilde{h} \in \tilde{U}_\alpha} (T(\tilde{h}) - I(\tilde{h})). \end{aligned} \quad (3.14)$$

Since  $\epsilon$  is arbitrary, this yields a chain of inequalities

$$\sup_{\tilde{h} \in \tilde{H}_{\alpha-\alpha^2}} (T(\tilde{h}) - I(\tilde{h})) \leq \sup_{\tilde{h} \in \tilde{U}_\alpha} (T(\tilde{h}) - I(\tilde{h})) \leq \liminf_{n \rightarrow \infty} \psi_{n,\alpha}^e \leq \limsup_{n \rightarrow \infty} \psi_{n,\alpha}^e \leq \sup_{\tilde{h} \in \tilde{H}_\alpha} (T(\tilde{h}) - I(\tilde{h})). \quad (3.15)$$

As  $\alpha \rightarrow 0+$ , the limits of  $\sup_{\tilde{H}_{\alpha-\alpha^2}} (T(\tilde{h}) - I(\tilde{h}))$  and  $\sup_{\tilde{H}_\alpha} (T(\tilde{h}) - I(\tilde{h}))$  are the same, so we have proven that

$$\psi^e = \lim_{\alpha \rightarrow 0} \lim_{n \rightarrow \infty} \psi_{n,\alpha}^e = \lim_{\alpha \rightarrow 0} \sup_{\tilde{h} \in \tilde{H}_\alpha} (T(\tilde{h}) - I(\tilde{h})). \quad (3.16)$$

By the compactness of  $\tilde{\mathcal{W}}$  and the continuity of  $e$ ,  $\tilde{H}_0 = \{\tilde{h} : e(\tilde{h}) = e\}$  is a nonempty compact set. We show that the right-hand side of (3.16) is equal to  $\sup_{\tilde{H}_0} (T(\tilde{h}) - I(\tilde{h}))$ . By definition, we can find a sequence of reduced graphons  $\tilde{h}_\alpha \in \tilde{H}_\alpha$  such that  $\lim_{\alpha \rightarrow 0} (T(\tilde{h}_\alpha) - I(\tilde{h}_\alpha)) = \limsup_{\tilde{H}_\alpha} (T(\tilde{h}) - I(\tilde{h}))$ . These reduced graphons converge to a reduced graphon  $\tilde{h}_0 \in \tilde{H}_0$ . Since  $T$  is continuous and  $I$  is lower semi-continuous,

$$\sup_{\tilde{H}_0} (T(\tilde{h}) - I(\tilde{h})) \geq T(\tilde{h}_0) - I(\tilde{h}_0) \geq \lim_{\alpha \rightarrow 0} (T(\tilde{h}_\alpha) - I(\tilde{h}_\alpha)). \quad (3.17)$$

However, since  $\tilde{H}_0 \subset \tilde{H}_\alpha$ ,  $\sup_{\tilde{H}_0} (T(\tilde{h}) - I(\tilde{h}))$  is at least as small as  $\sup_{\tilde{H}_\alpha} (T(\tilde{h}) - I(\tilde{h}))$ . Our claim thus follows.  $\square$

Fix  $e$ . Let  $\tilde{H}$  be the subset of  $\tilde{H}_0$  where  $T(\tilde{h}) - I(\tilde{h})$  is maximized. By the compactness of  $\tilde{H}_0$ , the continuity of  $T$  and the lower semi-continuity of  $I$ ,  $\tilde{H}$  is a nonempty compact set. Theorem 3.1 gives an asymptotic formula for  $\psi_{n,\alpha}^e$  but says nothing about the behavior of a typical random graph sampled from the constrained exponential model (3.2). In the unconstrained case (2.5) however, we know that the quotient image  $\tilde{h}^{G_n}$  of a sampled graph must lie close to the corresponding maximizing set  $\tilde{H}$  for  $\psi$  with probability vanishing in  $n$ . We expect that a similar phenomenon should occur in the constrained model as well, and this is confirmed by Theorem 3.2.

**Theorem 3.2.** *Take  $e \in [0, 1]$ . Let  $\tilde{H}$  be defined as above. Let  $\mathbb{P}_{n,\alpha}^e(G_n)$  (3.2) be the conditional probability measure on  $\mathcal{G}_n$ . Then for any  $\eta > 0$  and  $\alpha$  sufficiently small there exist  $C, \gamma > 0$  such that for all  $n$  large enough,*

$$\mathbb{P}_{n,\alpha}^e \left( \delta_\square(\tilde{h}^{G_n}, \tilde{H}) \geq \eta \right) \leq C e^{-n^2 \gamma}. \quad (3.18)$$

*Proof.* We check that the conditional probability measure  $\mathbb{P}_{n,\alpha}^e$  is well defined for all large enough  $n$ . It suffices to show that  $\psi_{n,\alpha}^e$  is finite. But from (3.15),  $\psi_{n,\alpha}^e$  is trapped between  $\sup_{\tilde{h} \in \tilde{U}_\alpha} (T(\tilde{h}) - I(\tilde{h}))$  and  $\sup_{\tilde{h} \in \tilde{H}_\alpha} (T(\tilde{h}) - I(\tilde{h}))$ , which are clearly both finite.

Recall that  $\tilde{H}_\alpha$  is the set of reduced graphons  $\tilde{h}$  with  $e - \alpha \leq e(\tilde{h}) \leq e + \alpha$ . Take any  $\eta > 0$ . Let  $\tilde{A}_\alpha$  be the subset of  $\tilde{H}_\alpha$  consisting of reduced graphons that are at least  $\eta$ -distance away from  $\tilde{H}$ ,

$$\tilde{A}_\alpha = \{\tilde{h} \in \tilde{H}_\alpha : \delta_\square(\tilde{h}, \tilde{H}) \geq \eta\}. \quad (3.19)$$

It is easy to see that  $\tilde{A}_\alpha$  is a closed set. Without loss of generality we assume that  $\tilde{A}_\alpha$  is nonempty for every  $\alpha > 0$ , since otherwise our claim trivially follows. Under this nonemptiness assumption we can find a sequence of reduced graphons  $\tilde{h}_\alpha \in \tilde{A}_\alpha$  converging to a reduced graphon  $\tilde{h}_0 \in \tilde{A}_0$ , which shows that  $\tilde{A}_0$  is nonempty as well. By the compactness of  $\tilde{H}_0$  and  $\tilde{H}$ ,

and the upper semi-continuity of  $T - I$ , it follows that

$$\max_{\tilde{h} \in \tilde{H}_0} (T(\tilde{h}) - I(\tilde{h})) - \max_{\tilde{h} \in \tilde{A}_0} (T(\tilde{h}) - I(\tilde{h})) > 0. \quad (3.20)$$

From the proof of Theorem 3.1 we see that

$$\limsup_{\tilde{H}_\alpha} (T(\tilde{h}) - I(\tilde{h})) = \max_{\tilde{H}_0} (T(\tilde{h}) - I(\tilde{h})). \quad (3.21)$$

Similarly, we have

$$\limsup_{\tilde{A}_\alpha} (T(\tilde{h}) - I(\tilde{h})) = \max_{\tilde{A}_0} (T(\tilde{h}) - I(\tilde{h})). \quad (3.22)$$

This implies that for  $\alpha$  sufficiently small,

$$2\gamma := \sup_{\tilde{h} \in \tilde{H}_{\alpha-\alpha^2}} (T(\tilde{h}) - I(\tilde{h})) - \sup_{\tilde{h} \in \tilde{A}_\alpha} (T(\tilde{h}) - I(\tilde{h})) > 0. \quad (3.23)$$

Choose  $\epsilon = \gamma$  and define  $\tilde{H}_{\alpha,c}$  and  $R$  as in the proof of Theorem 3.1. Let  $\tilde{A}_{\alpha,c} = \tilde{A}_\alpha \cap \tilde{H}_{\alpha,c}$ . Then

$$\mathbb{P}_{n,\alpha}^e(\tilde{h}^{G_n} \in \tilde{A}_\alpha) \leq \exp(-n^2 \psi_{n,\alpha}^\epsilon) |R| \sup_{c \in R} \exp(n^2(c + \gamma)) |\tilde{A}_{\alpha,c}^n|. \quad (3.24)$$

While bounding the last term above, it may be assumed without loss of generality that  $\tilde{A}_{\alpha,c}$  is nonempty for each  $c \in R$ . Similarly as in the proof of Theorem 3.1, the above inequality gives

$$\limsup_{n \rightarrow \infty} \frac{\log \mathbb{P}_{n,\alpha}^e(\tilde{h}^{G_n} \in \tilde{A}_\alpha)}{n^2} \leq \sup_{c \in R} \left( c + \gamma - \inf_{\tilde{h} \in \tilde{A}_{\alpha,c}} I(\tilde{h}) \right) - \sup_{\tilde{h} \in \tilde{H}_{\alpha-\alpha^2}} (T(\tilde{h}) - I(\tilde{h})). \quad (3.25)$$

Each  $\tilde{h} \in \tilde{A}_{\alpha,c}$  satisfies  $T(\tilde{h}) \geq c$ . Consequently,

$$\sup_{\tilde{h} \in \tilde{A}_{\alpha,c}} (T(\tilde{h}) - I(\tilde{h})) \geq c - \inf_{\tilde{h} \in \tilde{A}_{\alpha,c}} I(\tilde{h}). \quad (3.26)$$

Substituting this in (3.25) gives

$$\limsup_{n \rightarrow \infty} \frac{\log \mathbb{P}_{n,\alpha}^e(\tilde{h}^{G_n} \in \tilde{A}_\alpha)}{n^2} \leq \gamma + \sup_{\tilde{h} \in \tilde{A}_\alpha} (T(\tilde{h}) - I(\tilde{h})) - \sup_{\tilde{h} \in \tilde{H}_{\alpha-\alpha^2}} (T(\tilde{h}) - I(\tilde{h})) = -\gamma. \quad (3.27)$$

This completes the proof.  $\square$

#### 4. AN APPLICATION

Theorems 3.1 and 3.2 in the previous section illustrate the importance of finding the maximizing graphons for  $T - I$  subject to certain constraints. They aid us in understanding the limiting conditional probability distribution and the global structure of a random graph  $G_n$  drawn from the constrained exponential model. Indeed knowledge of such graphons would help us understand the limiting probability distribution and the global structure of a random graph  $G_n$  drawn from the unconstrained exponential model as well, since we can always carry out the unconstrained optimization in steps: first consider a constrained optimization (referred to as “micro analysis”), then take into consideration of all possible constraints (referred to as “macro analysis”). However, as straight-forward as it sounds, due to the myriad of structural possibilities of graphons, both the unconstrained (2.7) and constrained (3.4) optimization problems are not always explicitly solvable. So far major simplification has only been achieved in the “attractive” case [8, 19, 22] and for  $k$ -star models [8], whereas a complete analysis of either (2.7) or (3.4) in the “repulsive” region has proved to be very difficult. This section will provide

some phase transition results on the constrained “repulsive” edge-triangle exponential random graph model and discuss their possible generalizations. As we will see, they come in accordance with their unconstrained counterparts. We make these notions precise in the following.

The unconstrained edge-triangle model is a 2-parameter exponential random graph model obtained by taking  $H_1$  to be a single edge and  $H_2$  to be a triangle in (1.1). More explicitly, in the edge-triangle model, the probability measure  $\mathbb{P}_n^\beta$  is

$$\mathbb{P}_n^\beta(G_n) = \exp\left(n^2(\beta_1 e(G_n) + \beta_2 t(G_n) - \psi_n^\beta)\right), \quad (4.1)$$

where  $\beta = (\beta_1, \beta_2)$  are 2 real parameters,  $e(G_n)$  and  $t(G_n)$  are the edge and triangle densities of  $G_n$ , and  $\psi_n^\beta$  is the normalization constant. As before, we assume that the ideal edge density  $e$  is fixed. The limiting construction described at the beginning of Section 3 will then yield the asymptotic conditional normalization constant  $\psi^e$ . From (3.4) we see that  $\psi^e$  depends on both parameters  $\beta_1$  and  $\beta_2$ , however the  $\beta_1$  dependence is linear:  $\psi^e$  is equal to  $\beta_1 e$  plus a function independent of  $\beta_1$ . In particular  $\beta_1$  plays no role in the maximization problem, so we can consider it fixed at value  $\beta_1 = 0$ . The only relevant parameters then are  $e$  and  $\beta_2$ .

To highlight this parameter dependence, in the following we will write  $\psi^e$  as  $\psi^{e, \beta_2}$  instead. We are particularly interested in the asymptotics of  $\psi^{e, \beta_2}$  when  $\beta_2$  is negative, the so-called repulsive region. Naturally, varying  $\beta_2$  allows one to adjust the influence of the triangle density of the graph on the probability distribution. The more negative the  $\beta_2$ , the more unlikely that graphs with a large number of triangles will be observed. When  $\beta_2$  approaches negative infinity, the most probable graph would likely be triangle free. At the other extreme, when  $\beta_2$  is zero, the edge-triangle model reduces to the well-studied Erdős-Rényi model, where edges between different vertex pairs are independently included. The structure of triangle free graphs and disordered Erdős-Rényi graphs are apparently quite different, and thus a phase transition is expected as  $\beta_2$  decays from 0 to  $-\infty$ . In fact, it is believed that, quite generally, repulsive models exhibit a transition qualitatively like the solid/fluid transition, in that a region of parameter space depicting emergent multipartite structure, which is in imitation of the structure of solids, is separated by a phase transition from a region of disordered graphs, which resemble fluids. The existence of such a transition in unconstrained 2-parameter models whose subgraph  $H_2$  has chromatic number at least 3 has been proved by Aristoff and Radin [1] based on a symmetry breaking result from [8]. Theorem 4.1 below gives a corresponding result in the constrained edge-triangle model. Its proof though is quite different from the parallel result in [1] and relies instead on some analysis arguments. We also remark that, using the same arguments, it is possible to establish the phase transition as  $\beta_2$  grows from 0 to  $\infty$ , i.e., in the “attractive” region of the parameter space. There, combined with simulation results, we could conclude that a typical graph consists of one big clique and some isolated vertices as  $\beta_2$  gets sufficiently close to infinity.

**Theorem 4.1.** *Consider the constrained repulsive edge-triangle exponential random graph model as described above. Let  $e$  be arbitrary but fixed. Let  $\beta_2$  vary from 0 to  $-\infty$ . Then  $\psi^{e, \beta_2}$  is not analytic at at least one value of  $\beta_2$ .*

*Proof.* We first consider the case  $e \leq 1/2$ ; the case  $e > 1/2$  is similar, see the comments at the end of the proof.

Let  $e(\tilde{h}) \leq 1/2$  be the edge density of a reduced graphon  $\tilde{h}$  and  $t(\tilde{h})$  be the triangle density, obtained by taking  $H$  to be a triangle in (2.3). By (3.4),

$$\begin{aligned} \psi^{e,\beta_2} &= \sup_{\tilde{h} \in \tilde{\mathcal{W}}: e(\tilde{h})=e} \left( \beta_2 t(\tilde{h}) - I(\tilde{h}) \right) \\ &= \sup_t \sup_{\tilde{h} \in \tilde{\mathcal{W}}: e(\tilde{h})=e, t(\tilde{h})=t} \left( \beta_2 t - I(\tilde{h}) \right) \\ &= \sup_t (\beta_2 t + s(e, t)), \end{aligned} \quad (4.2)$$

where for notational convenience, we denote by  $s(e, t)$  the maximum value of  $-I(\tilde{h})$  over all reduced graphons with  $e(\tilde{h}) = e$  and  $t(\tilde{h}) = t$ . We examine (4.2) at the two extreme values of  $\beta_2$  first. Since  $I$  is convex, when  $\beta_2 = 0$ ,

$$\psi^{e,0} = \sup_{\tilde{h} \in \tilde{\mathcal{W}}: e(\tilde{h})=e} \left( -I(\tilde{h}) \right) \leq -I(e) \quad (4.3)$$

by Jensen's inequality, and the equality is attained only when  $h \equiv e$ , the associated graphon for an Erdős-Rényi graph with edge formation probability  $e$ . This also ensures that when we take  $\beta_2 \leq 0$ , any maximizing graphon  $h$  for (4.2) will satisfy  $t(\tilde{h}) \leq e^3$ . For the other extreme, take an arbitrary sequence  $\beta_2^{(i)} \rightarrow -\infty$ , and let  $\tilde{h}_i$  be a maximizing reduced graphon for each  $\psi^{e,\beta_2^{(i)}}$ . Let  $\tilde{h}$  be a limit point of  $\tilde{h}_i$  in  $\tilde{\mathcal{W}}$  (its existence is guaranteed by the compactness of  $\tilde{\mathcal{W}}$ ). Suppose  $t(\tilde{h}) > 0$ . Then by the continuity of  $t$  and the boundedness of  $I$ ,  $\lim_{i \rightarrow \infty} \psi^{e,\beta_2^{(i)}} = -\infty$ . But this is impossible since  $\psi^{e,\beta_2^{(i)}}$  is uniformly bounded below, as can be seen by considering

$$h(x, y) = \begin{cases} 2e & \text{if } x < 1/2 < y \text{ or } x > 1/2 > y; \\ 0 & \text{if } x, y < 1/2 \text{ or } x, y > 1/2, \end{cases} \quad (4.4)$$

as a test function. Thus  $t(\tilde{h}) = 0$ . The rest of the proof will utilize the following useful features of  $s(e, t)$  derived in Radin and Sadun [17, 18]: for  $e \leq 1/2$ ,  $s(e, 0) = -I(2e)/2$  and this maximum is achieved only at the reduced graphon  $\tilde{h}$  (4.4), corresponding to a complete bipartite graph with  $1 - 2e$  fraction of edges randomly deleted. Moreover, for any  $e \in [0, 1]$  and for  $t \leq e^3$ ,

$$s(e, e^3) - s(e, t) \geq c(e^3 - t)^{2/3} \quad (4.5)$$

for some  $c = c(e) > 0$ . Thus we have

$$\lim_{\beta_2 \rightarrow -\infty} \psi^{e,\beta_2} = -I(2e)/2; \quad (4.6)$$

while (4.5) implies that for  $\beta_2 > -c(e)$  and  $t < e^3$ ,

$$-\beta_2(e^3 - t) < s(e, e^3) - s(e, t). \quad (4.7)$$

In other words, the constant graphon  $h \equiv e$  still yields the maximum value for (4.2) for these small values of  $\beta_2$ . Thus regarded as a function of  $\beta_2$ ,  $\psi^{e,\beta_2}$  is constant on the interval  $(-c(e), 0)$  and  $\psi^{e,\beta_2} = -I(e)$ . This shows that  $\psi^{e,\beta_2}$  must lose its analyticity at at least one  $\beta_2$  as  $\beta_2$  varies from 0 to  $-\infty$ , since otherwise we would have

$$\lim_{\beta_2 \rightarrow -\infty} \psi^{e,\beta_2} = -I(e), \quad (4.8)$$

in contradiction with (4.6).

For  $e > 1/2$ , the lower boundary of attainable  $t(\tilde{h})$  is nonzero; see Figure 1. However the graphons attaining the minimum  $t$  values for each  $e$  are known, see [17], and their rate functions are strictly less than  $-I(e)$ , so the proof above goes through without change.  $\square$

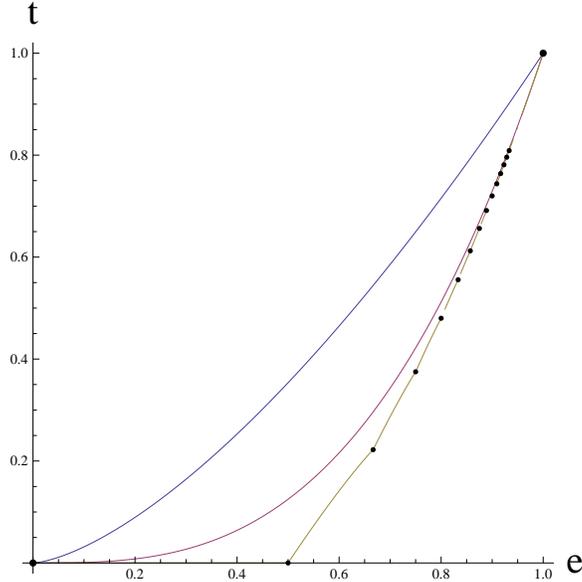


FIGURE 1. Region of attainable edge ( $e$ ) and triangle ( $t$ ) densities for graphons. The upper boundary is the curve  $t = e^{3/2}$  and the lower boundary is a piecewise algebraic curve with infinitely many concave pieces; see [20]. The red curve is the Erdős-Rényi curve  $t = e^3$ .

The proof of Theorem 4.1 does not rely heavily on the definition of the edge-triangle model, except for the non-differentiability of  $s(e, t)$  at  $t = e^3$  and the structure of the maximizing graphons at the two extreme values of  $\beta_2$ . The following extension of this theorem may not come as a surprise.

**Theorem 4.2.** *Take  $H_1$  a single edge and  $H_2$  a different, arbitrary simple graph with chromatic number  $\chi(H_2)$  at least 3. Consider the constrained repulsive 2-parameter exponential random graph model where the probability measure  $\mathbb{P}_n^{e, \beta_2}$  is given by*

$$\mathbb{P}_n^{e, \beta_2}(G_n) = \exp\left(n^2(\beta_2 t(H_2, G_n) - \psi_n^{e, \beta_2})\right). \quad (4.9)$$

*Let the edge density  $e$  be fixed. Let the second parameter  $\beta_2$  vary from 0 to  $-\infty$ . Then  $\psi^{e, \beta_2}$  loses its analyticity at at least one value of  $\beta_2$ .*

*Proof.* The proof of Theorem 4.1 carries over almost word-for-word when we incorporate the disordered Erdős-Rényi structure of the maximizing graphon at  $\beta_2 = 0$ , the non-differentiability of  $s(e, t)$  for a general  $H_2$  [18], and the emergent multipartite structure of the maximizing graphon as  $\beta_2 \rightarrow -\infty$  [8, 23].  $\square$

Now that we know about the occurrence of a phase transition in the constrained repulsive exponential model, we probe deeper into this phenomenon and ask: how smooth is this transition? Theorem 4.3 shows what happens when the ideal edge density of the edge-triangle model is fixed at  $1/2$  while the influence of the triangle density is tuned through the parameter  $\beta_2$ .

**Theorem 4.3.** *Consider the constrained repulsive edge-triangle exponential random graph model as described at the beginning of Section 4. Fix  $e = 1/2$ . Let  $\beta_2$  vary from 0 to  $-\infty$ . Then  $\psi^{e,\beta_2}$  is analytic everywhere except at a certain point  $\beta_2^c$ , where the derivative  $\frac{\partial}{\partial \beta_2} \psi^{e,\beta_2}$  displays jump discontinuity.*

*Proof.* Setting  $e = 1/2$  in (4.2) gives

$$\psi^{\frac{1}{2},\beta_2} = \sup_t (\beta_2 t + s(\frac{1}{2}, t)). \quad (4.10)$$

Since  $\beta_2 \leq 0$ , by the convexity of  $I$ , any maximizing graphon  $h$  for (4.10) must satisfy  $t(\tilde{h}) \leq 1/8$ , i.e., it must lie below the Erdős-Rényi curve  $t = e^3$ . Radin and Sadun [18] showed that on the line segment  $e = \frac{1}{2}$  and  $t \leq e^3$ , the symmetric bipodal graphon

$$h(x, y) = \begin{cases} \frac{1}{2} + \epsilon, & \text{if } x < \frac{1}{2} < y \text{ or } x > \frac{1}{2} > y; \\ \frac{1}{2} - \epsilon, & \text{if } x, y < \frac{1}{2} \text{ or } x, y > \frac{1}{2}, \end{cases} \quad (4.11)$$

where  $0 \leq \epsilon = (\frac{1}{8} - t)^{\frac{1}{3}} \leq \frac{1}{2}$ , maximizes  $s(\frac{1}{2}, t)$ , and that every maximizing graphon is of the form  $h_\sigma$  for some measure preserving bijection  $\sigma$ . Equivalently, the maximum value for (4.10) is achieved only at the reduced bipodal graphon  $\tilde{h}$ . See Figure 2 for the graph of  $s(1/2, t)$ .

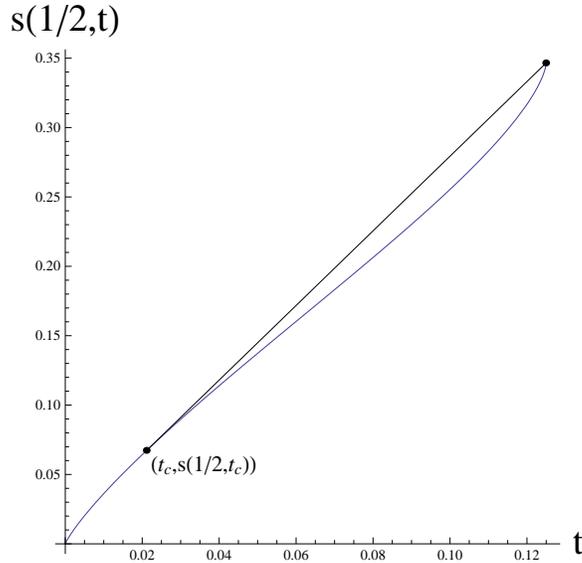


FIGURE 2. The graph of  $s(1/2, t)$  below the ER line for the edge-triangle model (blue). The convex hull of the region below the graph is delimited by the black line segment and the portion of the graph to its left; this segment is the support line at the right endpoint ( $t = 1/8, s = \frac{\log 2}{2}$ ) of maximal slope  $-\beta_2^c$ . The other point at which the line segment meets the curve is the point  $(t_c, s(1/2, t_c))$ .

Geometrically, the maximization problem in (4.10) involves finding the lowest half-plane with bounding line of slope  $-\beta_2$  lying above the graph of  $s(1/2, t)$ . For  $\beta_2 > \beta_2^c$  the boundary of this half-plane passes only through the graph of  $s(1/2, t)$  at the right endpoint  $(1/8, \frac{\log 2}{2})$ . The critical value  $\beta_2^c$  is defined (as in the Figure) as the first slope at which this half-plane intersects

the curve at a different point. We let  $(t_c, s(1/2, t_c))$  be this second point. At more negative values of  $\beta_2$ , the half-plane will hit the curve at points with  $t$  values below  $t_c$ .

In particular this shows the non-analyticity of  $\psi^{\frac{1}{2}, \beta_2}$  as a function of  $\beta_2$  at  $\beta_2 = \beta_2^c$ . The analyticity of  $\psi^{\frac{1}{2}, \beta_2}$  elsewhere follows from concavity (and analyticity) of  $s(1/2, t)$  below  $t_c$ . By Theorem 3.2, at  $\beta_2 = \beta_2^c$ , the maximizing reduced graphon  $\tilde{h}$  for (4.10) transitions from being Erdős-Rényi with edge formation probability  $\frac{1}{2}$  to symmetric bipodal with  $\epsilon_c = (\frac{1}{8} - t_c)^{1/3}$ . The jump discontinuity in the derivative follows when we realize that  $\frac{\partial}{\partial \beta_2} \psi^{\frac{1}{2}, \beta_2} = t(\tilde{h})$ .  $\square$

Numerical computations yield that  $\beta_2^c$  is approximately  $-2.7$  and  $\epsilon_c$  is approximately  $0.47$ . By Theorem 3.2, this shows that as  $\beta_2$  decreases from  $0$  to  $-\infty$ , a typical graph  $G_n$  drawn from the constrained repulsive edge-triangle model jumps from being Erdős-Rényi to almost complete bipartite, skipping a large portion of the  $e = \frac{1}{2}$  line. This “jump behavior” (also called first-order phase transition) is intrinsically tied to the *convexity* of  $s(e, t)$  just below the Erdős-Rényi curve  $t = e^3$ , thus we expect similar phase transition phenomena for general  $e \neq \frac{1}{2}$  as well; see Figures 3, 4. See [8] for related results in the unconstrained repulsive edge-triangle model.

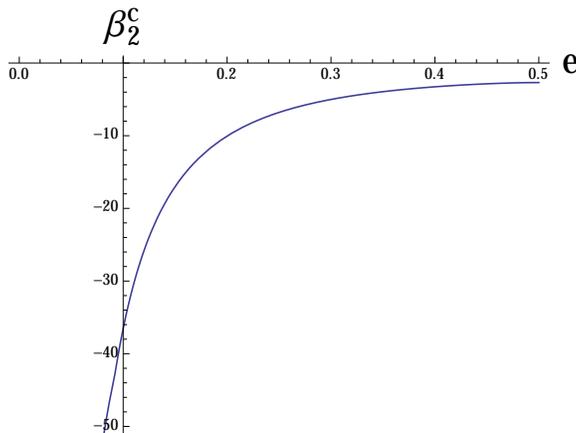


FIGURE 3. The (conjectural) graph of  $\beta_2^c$  as a function of  $e$  for the edge-triangle model, in the range  $e \leq 1/2$ . The computation is based on the conjecture that the maximizing graphons in this region are symmetric and bipodal, see [16].

## 5. EULER-LAGRANGE EQUATIONS

We return to the constrained 2-parameter family of exponential random graphs (4.9). For notational convenience and with some abuse of notation, denote by  $T(h) = \sum_{i=1}^2 \beta_i t(H_i, h)$ . As seen in Section 4, the “micro analysis” helps with the “macro analysis”. Explicitly, if we can find the maximizing graphon for  $-I$  subject to two constraints  $t(H_1, \cdot) = t_1$  and  $t(H_2, \cdot) = t_2$ , where  $t_1$  and  $t_2$  are arbitrary but fixed homomorphism densities, then we can find the maximizing graphon for  $T - I$  subject to fewer or even no constraints. This in turn will aid us in understanding the limiting conditional probability distribution and the structure of a typical graph  $G_n$  sampled from either the constrained or the unconstrained exponential model. In the unconstrained case, Chatterjee and Diaconis derived the Euler-Lagrange equation for

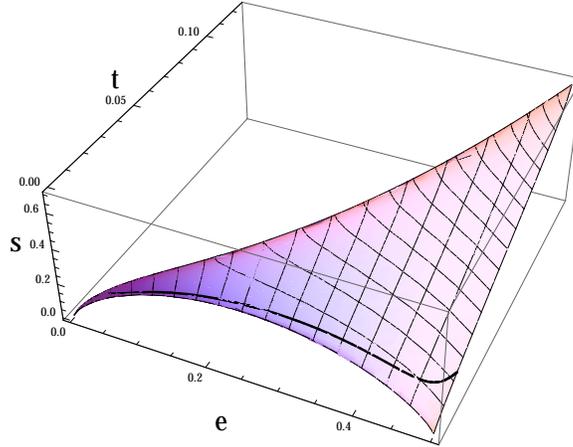


FIGURE 4. The (conjectural) graph of entropy  $-I$  as a function of edge and triangle densities  $e, t$  in the region  $e \leq \frac{1}{2}, t \leq e^3$ . The critical curve (black) defines  $t^c$  as a function of  $e$ . The computation is based on the conjecture that the maximizing graphons in this region are symmetric and bipodal, see [16].

the maximizing graphon  $h$  for  $T(h) - I(h)$  when the tuning parameters are arbitrary but fixed (Theorem 6.1 in [8]). When applied to the 2-parameter model, they showed that

$$h(x, y) = \frac{e^{2 \sum_{i=1}^2 \beta_i \Delta_{H_i} h(x, y)}}{1 + e^{2 \sum_{i=1}^2 \beta_i \Delta_{H_i} h(x, y)}}, \quad (5.1)$$

where for a finite simple graph  $H$  with vertex set  $V(H)$  and edge set  $E(H)$ ,

$$\Delta_H h(x, y) = \sum_{(r, s) \in E(H)} \Delta_{H, r, s} h(x, y), \quad (5.2)$$

and for each  $(r, s) \in E(H)$  and each pair of points  $x_r, x_s \in [0, 1]$ ,

$$\Delta_{H, r, s} h(x_r, x_s) = \int_{[0, 1]^{|V(H) \setminus \{r, s\}|}} \prod_{(r', s') \in E(H): (r', s') \neq (r, s)} h(x_{r'}, x_{s'}) \prod_{v \in V(H): v \neq r, s} dx_v. \quad (5.3)$$

For example, in the edge-triangle model where  $H_1$  is an edge and  $H_2$  is a triangle,  $\Delta_{H_1} h(x, y) \equiv 1$  and  $\Delta_{H_2} h(x, y) = 3 \int_0^1 h(x, z) h(y, z) dz$ . In the constrained case, we could likewise derive the Euler-Lagrange equation by resorting to the method of Lagrange multipliers, which will turn the constrained maximization into an unconstrained one, but we provide an alternative bare-hands approach here. The following theorem may also be formulated in terms of reduced graphons.

**Theorem 5.1.** *Consider the constrained 2-parameter exponential random graph model (4.9). Let  $t_1$  and  $t_2$  be arbitrary but fixed homomorphism densities. Suppose the graphon  $h$  maximizes  $-I(h)$  subject to  $t(H_1, h) = t_1$  and  $t(H_2, h) = t_2$ . If  $h$  is bounded away from 0 and 1, then there must exist constants  $\beta_1$  and  $\beta_2$  such that  $h$  satisfies (5.1) for almost all  $(x, y) \in [0, 1]^2$ .*

*Proof.* Graphons are bounded integrable functions on  $[0, 1]^2$  so they are almost everywhere continuous. Let  $(x_i, y_i)$  for  $i = 1, 2, 3$  be three points of  $[0, 1]^2$  and let  $h_i = h(x_i, y_i)$ . Without loss of generality we assume that  $h_1 \neq h_2$ , since otherwise  $h$  is a constant graphon and our

claim trivially follows. We infinitesimally perturb the values at these three points, sending  $h_i \rightarrow h_i + dh_i$ . Then  $(t_1, t_2, -I) \rightarrow (t_1, t_2, -I) + (dt_1, dt_2, -dI)$  where

$$\begin{pmatrix} dt_1 \\ dt_2 \\ -dI \end{pmatrix} = \begin{pmatrix} \Delta_{H_1} h_1 & \Delta_{H_1} h_2 & \Delta_{H_1} h_3 \\ \Delta_{H_2} h_1 & \Delta_{H_2} h_2 & \Delta_{H_2} h_3 \\ \frac{1}{2} \log(\frac{1}{h_1} - 1) & \frac{1}{2} \log(\frac{1}{h_2} - 1) & \frac{1}{2} \log(\frac{1}{h_3} - 1) \end{pmatrix} \begin{pmatrix} dh_1 \\ dh_2 \\ dh_3 \end{pmatrix}. \quad (5.4)$$

If the determinant of the above matrix is nonzero, then there is a nontrivial deformation  $(dh_1, dh_2, dh_3)$  which increases  $-I$  while leaving  $t_1$  and  $t_2$  fixed. So the maximizing graphon  $h$  must satisfy the condition that the determinant is zero. Recall that  $H_1$  is a single edge and  $\Delta_{H_1} h_i \equiv 1$ . Thus the first and third rows of the matrix are linearly independent and there must exist constants  $\beta_1$  and  $\beta_2$  such that

$$\Delta_{H_2} h_i = \beta_1 + \frac{\beta_2}{2} \log(\frac{1}{h_i} - 1). \quad (5.5)$$

Moreover, since  $\beta_1$  and  $\beta_2$  are determined by  $h_1$  and  $h_2$ , we must have (5.5) for all points  $(x_3, y_3) \in [0, 1]^2$ . We recognize this requirement is equivalent to (5.1).  $\square$

Suppose we are looking for a graphon  $h$  that maximizes  $-I(h)$  subject to  $t(H_1, h) = t_1$  only. Then following the same ‘‘perturbation’’ idea, we should examine

$$\begin{pmatrix} dt_1 \\ -dI \end{pmatrix} = \begin{pmatrix} \Delta_{H_1} h_1 & \Delta_{H_1} h_2 \\ \frac{1}{2} \log(\frac{1}{h_1} - 1) & \frac{1}{2} \log(\frac{1}{h_2} - 1) \end{pmatrix} \begin{pmatrix} dh_1 \\ dh_2 \end{pmatrix}. \quad (5.6)$$

Since the determinant is zero,  $h$  must be a constant. This is the same conclusion obtained by applying Jensen’s inequality to the convex function  $I$ . On the other hand, we may also consider maximizing  $-I(h)$  subject to  $k$  (instead of 2) constraints  $t(H_i, h) = t_i$  for  $i = 1, \dots, k$ , in which case we would perturb the values of the graphon at  $k + 1$  points and form a  $(k + 1) \times (k + 1)$  matrix.

## REFERENCES

- [1] Aristoff, D, Radin, C.: Emergent structures in large networks. *J. Appl. Prob.* 50, 883-888 (2013)
- [2] Aldous, D.: Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* 11, 581-598 (1981)
- [3] Bollobás, B.: *Random Graphs*, Volume 73 of Cambridge Studies in Advanced Mathematics. 2nd ed. Cambridge University Press, Cambridge (2001)
- [4] Borgs, C., Chayes, J.T., Cohn, H., Zhao, Y: An  $L^p$  theory of sparse graph convergence I. Limits, sparse random graph models, and power law distributions. arXiv: 1401.2906 (2014)
- [5] 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)
- [6] 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)
- [7] 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)
- [8] Chatterjee, S., Diaconis, P.: Estimating and understanding exponential random graph models. *Ann. Statist.* 41, 2428-2461 (2013)
- [9] Chatterjee, S., Diaconis, P., Sly, A.: Random graphs with a given degree sequence. *Ann. Appl. Prob.* 21, 1400-1435 (2011)
- [10] Chatterjee, S., Varadhan, S.R.S.: The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.* 32, 1000-1017 (2011)
- [11] Frank, O., Strauss, D.: Markov graphs. *J. Amer. Statist. Assoc.* 81, 832-842 (1986)
- [12] 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)

- [13] Lovász, L.: Large Networks and Graph Limits. American Mathematical Society, Providence (2012)
- [14] Lovász, L., Szegedy B.: Limits of dense graph sequences. J. Combin. Theory Ser. B 96, 933-957 (2006)
- [15] Newman, M.: Networks: An Introduction. Oxford University Press, New York (2010)
- [16] Radin, C., Ren, K., Sadun, L.: The asymptotics of large constrained graphs. arXiv: 1401.1170 (2014)
- [17] Radin, C., Sadun, L.: Phase transitions in a complex network. J. Phys. A 46, 305002 (2013)
- [18] Radin, C., Sadun, L.: Singularities in the entropy of asymptotically large simple graphs. arXiv: 1302.3531 (2013)
- [19] Radin, C., Yin, M.: Phase transitions in exponential random graphs. Ann. Appl. Probab. 23, 2458-2471 (2013)
- [20] Razborov, A.: On the minimal density of triangles in graphs. Combin. Probab. Comput. 17, 603-618 (2008)
- [21] Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (2010)
- [22] Yin, M.: Critical phenomena in exponential random graphs. J. Stat. Phys. 153, 1008-1021 (2013)
- [23] Yin, M., Rinaldo, A., Fadnavis, S.: Asymptotic quantization of exponential random graphs. arXiv: 1311.1738 (2013)

DEPARTMENT OF MATHEMATICS, BROWN UNIVERSITY, PROVIDENCE, RI 02912, USA

*E-mail address:* rkenyon at math.brown.edu, mei\_yin at math.brown.edu