J. Appl. Prob. **54,** 165–180 (2017)
*doi:*10.1017/jpr.2016.93
© *Applied Probability Trust* 2017

ON THE ASYMPTOTICS OF CONSTRAINED EXPONENTIAL RANDOM GRAPHS

RICHARD KENYON,* *Brown University*
MEI YIN,** *University of Denver*

Abstract

The unconstrained exponential family of random graphs assumes no prior knowledge of the graph before sampling, but it is natural to consider situations where partial information about the graph is known, for example the total number of edges. What does a typical random graph look like, if drawn from an exponential model subject to such constraints? Will there be a similar phase transition phenomenon (as one varies the parameters) as that which occurs in the unconstrained exponential model? We present some general results for this constrained model and then apply them to obtain concrete answers in the edge-triangle model with fixed density of edges.

Keywords: Constrained exponential random graph; phase transition

2010 Mathematics Subject Classification: Primary 05C80
Secondary 82B26

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 $G_n \in \mathcal{G}_n$,

$$\mathbb{P}_n^\beta(G_n) = \exp(n^2(\beta_1 t(H_1, G_n) + \cdots + \beta_k t(H_k, G_n) - \psi_n^\beta)), \tag{1}$$

where $\beta = (\beta_1, \ldots, \beta_k)$ are k real parameters, $H_1, \ldots, H_k$ are prechosen 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 $V(H_i) \to V(G_n)$ is edge-preserving),

$$t(H_i, G_n) = \frac{|\hom(H_i, G_n)|}{|V(G_n)|^{|V(H_i)|}}, \tag{2}$$

and ψ_n^β is the normalization constant,

$$\psi_n^\beta = \frac{1}{n^2} \log \sum_{G_n \in \mathcal{G}_n} \exp(n^2(\beta_1 t(H_1, G_n) + \cdots + \beta_k t(H_k, G_n))). \tag{3}$$

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.

Received 29 September 2015; revision received 22 June 2016.
* Postal address: Department of Mathematics, Brown University, Providence, RI 02912, USA.
Email address: rkenyon@math.brown.edu
** Postal address: Department of Mathematics, University of Denver, Denver, CO 80208, USA.
Email address: mei.yin@du.edu

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

Exponential random graphs have been 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], [12], [17], [24], [25]. Intuitively, we can think of the k-parameters $\beta_1, \ldots, \beta_k$ as tuning parameters that allow one to adjust the influence of different subgraphs $H_1, \ldots, H_k$ of G_n on the probability distribution, whose asymptotics are our main interest since networks are often very large in size. As flexible as they are, exponential models admittedly have one shortcoming: they are centered on dense graphs whereas most network data in the real world are sparse. In this sense, one could argue that exponential random graphs (and the graphon technology developed by Lovász *et al.* [5]–[7], [14], [15] that is heavily used in studying them) are of limited relevance in studying real networks. However, from the point of view of extremal combinatorics and statistical mechanics, exponential random graphs and constrained graphons represent an important and challenging class of models, displaying both diverse and novel phase transition behavior [18]–[21].

Our main results are (Theorem 1) a variational principle for the normalization constant (partition function) for graphons with constrained edge density, and an associated concentration of measure (Theorem 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.

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 [1] and Hoover [13], Lovász and coauthors constructed an elegant theory of graph limits in a sequence of papers, [5]–[7], [15]. See also the recent book [14] 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

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

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 $V(H) = [k] = \{1, \ldots, k\}$ and edge set $E(H)$,

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

where $t(H, h^{G_n}) = t(H, G_n)$, the graph homomorphism density (2), by construction, and

$$t(H, h) = \int_{[0,1]^k} \prod_{\{i,j\} \in E(H)} h(x_i, x_j) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k. \tag{4}$$

Indeed, every function in $\mathcal{W}$ is the limit of a certain convergent graph sequence [15]. Intuitively, the interval $[0, 1]$ represents a 'continuum' of vertices, and $h(x, y)$ denotes the probability of

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

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 ρ 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

$$d_{\square}(f, h) = \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} (f(x, y) - h(x, y)) \, \mathrm{d}x \, \mathrm{d}y \right| \quad \text{for } f, h \in \mathcal{W}.$$

A nontrivial 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 σ 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 σ, 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 σ_1 and σ_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 $T\colon \tilde{\mathcal{W}} \to \mathbb{R}$ be a bounded continuous function. Let the probability measure $\mathbb{P}_n$ and the normalization constant ψ_n be defined as in (1) and (3), that is,

$$\begin{aligned} \mathbb{P}_n(G_n) &= \exp(n^2(T(\tilde{h}^{G_n}) - \psi_n)), \\ \psi_n &= \frac{1}{n^2} \log \sum_{G_n\in\mathcal{G}_n} \exp(n^2 T(\tilde{h}^{G_n})). \end{aligned} \tag{5}$$

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

$$\psi = \sup_{\tilde{h}\in\tilde{\mathcal{W}}} (T(\tilde{h}) - I(\tilde{h})), \tag{6}$$

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

$$I(u) = \tfrac{1}{2}u \log u + \tfrac{1}{2}(1 - u) \log(1 - u), \tag{7}$$

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

$$I(\tilde{h}) = \int_{[0,1]^2} I(h(x, y)) \, \mathrm{d}x \, \mathrm{d}y, \tag{8}$$

where h is any representative element of the equivalence class $\tilde{h}$. It was shown in [10] that I is well defined and lower semicontinuous on $\tilde{\mathcal{W}}$. Let $\tilde{H}$ be the subset of $\tilde{\mathcal{W}}$ where ψ is

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

maximized. By the compactness of $\tilde{\mathcal{W}}$, the continuity of T and the lower semicontinuity of I, $\tilde{H}$ is a nonempty compact set. The set $\tilde{H}$ encodes important information about the exponential model (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 (5) must lie close to $\tilde{H}$ with high probability,

$$\delta_{\square}(\tilde{h}^{G_n}, \tilde{H}) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty, \tag{9}$$

where '$\xrightarrow{\mathbb{P}}$' denotes convergence in probability. Since the limiting normalization constant ψ 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 ψ is nonanalytic or when $\tilde{H}$ is not a singleton set.

3. Constrained exponential random graphs

The exponential family of random graphs introduced above have popular counterparts in statistical physics: a hierarchy of models ranging from the grand canonical ensemble, the canonical ensemble, to the microcanonical ensemble with subgraph densities in place of particle and energy densities, and tuning parameters in place of temperature and chemical potentials. In the grand canonical ensemble, the exponential model (1) in this case, no prior knowledge of the graph is assumed. As useful as they are, for large networks these models are sometimes inappropriate. For example, as shown by Chatterjee and Diaconis [8], when $k = 2$ and $\beta_2 > 0$, all graphs drawn from (1) where H_1 is an edge and H_2 is any finite simple graph are not appreciably different from Erdős–Rényi in the large n limit. This somewhat trivial conclusion implies that sometimes subgraph densities cannot be tuned and exponential random graphs alone may not capture all the desirable features of the networked system, such as interdependency and clustering. We are thus motivated to study variants of the exponential random graph model: the canonical ensemble, where some subgraph density is controlled directly and others are tuned with parameters, and the microcanonical ensemble, where complete information of the graph is observed beforehand.

One difficulty arises. Unlike standard statistical physics models, the equivalence of various ensembles in the asymptotic regime does not hold in these models (see [23] for discussions about nonequivalence of ensembles due to nonconcavity of entropy). A natural question to ask is what would be a typical random graph drawn from an exponential model subject to certain constraints? Or, perhaps more importantly, will there be a similar phase transition phenomenon as in the standard exponential model (hereby referred to as an unconstrained model)? The following Theorems 1 and 2 provide 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, there are noted differences in how we interpret these phase transition results. For example, a typical graph drawn from the constrained edge-triangle model still exhibits Erdős–Rényi structure for β_2 close to 0, but consists of one big clique and some isolated vertices as β_2 gets sufficiently close to ∞, so the transition is between graphs of different characters. In the unconstrained model, on the other hand, although there is a curve in the parameter space across which the graph densities display sudden jumps [8], [20], the transition is between graphs of similar characters (Erdős–Rényi graphs). This gives one more reason why the constrained model deserves its own attention. Due to the imposed constraints, instead of working with probability measure $\mathbb{P}_n$ and normalization constant ψ_n as in [8], we are working with a conditional probability measure and a conditional normalization constant, so the argument is more involved. The proof of Theorem 1 also incorporates some ideas from Theorem 3.1 of [18].

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

For clarity, we assume that the edge density of the graph is approximately known, though the proof runs through without much modification if the density of some other more complicated subgraph is approximately described. We make precise the notion of 'approximately' below. We still assign a probability measure $\mathbb{P}_n$ as in (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 $\psi^e_{n,\alpha}$ is defined analogously to the normalization constant for the unconstrained exponential random graph model,

$$\psi^e_{n,\alpha} = \frac{1}{n^2} \log \sum_{\{G_n \in \mathcal{G}_n \colon |e(G_n)-e|<\alpha\}} \exp(n^2 T(\tilde{h}^{G_n})), \tag{10}$$

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

$$\mathbb{P}^e_{n,\alpha}(G_n) = \exp(-n^2 \psi^e_{n,\alpha}) \exp(n^2 T(\tilde{h}^{G_n})) \mathbf{1}_{\{|e(G_n)-e|<\alpha\}}, \tag{11}$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator. We perform two limit operations on $\psi^e_{n,\alpha}$. First we take n to ∞, then we shrink the interval around e by letting α go to 0, that is,

$$\psi^e = \lim_{\alpha \to 0} \lim_{n \to \infty} \psi^e_{n,\alpha}. \tag{12}$$

Intuitively, these two operations ensure that we are examining the asymptotics of exponentially weighted large graphs with edge density sufficiently close to e. In Theorem 1 we show that this is indeed the case.

Theorem 1. *Let $e \colon 0 \le e \le 1$ be a real parameter and $T \colon \tilde{\mathcal{W}} \to \mathbb{R}$ be a bounded continuous function. Let I and ψ^e be defined as before (see (7), (8), (10), and (12)). Then*

$$\psi^e = \sup_{\{\tilde{h} \in \tilde{\mathcal{W}} \colon e(\tilde{h})=e\}} (T(\tilde{h}) - I(\tilde{h})), \tag{13}$$

where

$$e(\tilde{h}) = \int_{[0,1]^2} h(x_1, x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2,$$

and h is any function in the equivalence class $\tilde{h}$.

Proof. By definition, $\liminf \psi^e_{n,\alpha}$ and $\limsup \psi^e_{n,\alpha}$ exist as $n \to \infty$. We will show that they both approach $\sup_{\{\tilde{h} \colon e(\tilde{h})=e\}}(T(\tilde{h}) - I(\tilde{h}))$ as $\alpha \to 0$. 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 $e - \alpha \le e(\tilde{h}) \le e + \alpha$. Fix $\varepsilon > 0$. Since T is a bounded function, there is a finite set R such that the intervals $\{(c, c + \varepsilon) \colon c \in R\}$ 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 $c < T(\tilde{h}) < c + \varepsilon$, and let $\tilde{H}_{\alpha,c}$ be the closed set $e - \alpha \le e(\tilde{h}) \le e + \alpha$ and $c \le T(\tilde{h}) \le c + \varepsilon$. 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 $|\tilde{U}^n_{\alpha,c}|$ and $|\tilde{H}^n_{\alpha,c}|$ 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

$$\limsup_{n \to \infty} \frac{\log |\tilde{H}^n_{\alpha,c}|}{n^2} \le - \inf_{\tilde{h} \in \tilde{H}_{\alpha,c}} I(\tilde{h}),$$

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

and that

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

We first consider $\limsup \psi^e_{n,\alpha}$. Thus,

$$\exp(n^2\psi^e_{n,\alpha}) \leq \sum_{c\in R} \exp(n^2(c+\varepsilon))|\tilde{H}^n_{\alpha,c}| \leq |R| \sup_{c\in R} \exp(n^2(c+\varepsilon))|\tilde{H}^n_{\alpha,c}|.$$

This shows that

$$\limsup_{n\to\infty} \psi^e_{n,\alpha} \leq \sup_{c\in R}\Big(c+\varepsilon - \inf_{\tilde{h}\in\tilde{H}_{\alpha,c}} I(\tilde{h})\Big). \tag{14}$$

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

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

Substituting this in (14), we have

$$\limsup_{n\to\infty} \psi^e_{n,\alpha} \leq \varepsilon + \sup_{c\in R} \sup_{\tilde{h}\in\tilde{H}_{\alpha,c}} (T(\tilde{h}) - I(\tilde{h})) = \varepsilon + \sup_{\tilde{h}\in\tilde{H}_\alpha} (T(\tilde{h}) - I(\tilde{h})).$$

Next we consider $\liminf \psi^e_{n,\alpha}$. Thus,

$$\exp(n^2\psi^e_{n,\alpha}) \geq \sup_{c\in R} \exp(n^2 c)|\tilde{U}^n_{\alpha,c}|.$$

Therefore, for each $c \in R$,

$$\liminf_{n\to\infty} \psi^e_{n,\alpha} \geq c - \inf_{\tilde{h}\in\tilde{U}_{\alpha,c}} I(\tilde{h}). \tag{15}$$

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

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

Together with (15), we show that

$$\liminf_{n\to\infty} \psi^e_{n,\alpha} \geq -\varepsilon + \sup_{c\in R} \sup_{\tilde{h}\in\tilde{U}_{\alpha,c}} (T(\tilde{h}) - I(\tilde{h})) = -\varepsilon + \sup_{\tilde{h}\in\tilde{U}_\alpha} (T(\tilde{h}) - I(\tilde{h})).$$

Since ε is arbitrary, this yields a chain of inequalities

$$\begin{aligned}
\sup_{\tilde{h}\in\tilde{H}_{\alpha-\alpha^2}} (T(\tilde{h}) - I(\tilde{h})) &\leq \sup_{\tilde{h}\in\tilde{U}_\alpha} (T(\tilde{h}) - I(\tilde{h})) \\
&\leq \liminf_{n\to\infty} \psi^e_{n,\alpha} \\
&\leq \limsup_{n\to\infty} \psi^e_{n,\alpha} \\
&\leq \sup_{\tilde{h}\in\tilde{H}_\alpha} (T(\tilde{h}) - I(\tilde{h})).
\end{aligned} \tag{16}$$

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

As $\alpha \to 0+$, 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

$$\psi^e = \lim_{\alpha\to 0}\lim_{n\to\infty}\psi^e_{n,\alpha} = \lim_{\alpha\to 0}\sup_{\tilde{h}\in\tilde{H}_\alpha}(T(\tilde{h}) - I(\tilde{h})). \tag{17}$$

First we establish that the right-hand side of (17) is equal to $\sup_{\tilde{H}_0}(T(\tilde{h}) - I(\tilde{h}))$, where $\tilde{H}_0 = \{\tilde{h}\colon e(\tilde{h}) = e\}$. By the compactness of $\tilde{\mathcal{W}}$ and the continuity of e, $\tilde{H}_0$ is a nonempty compact set. By definition, we can find a sequence of reduced graphons $\tilde{h}_\alpha \in \tilde{H}_\alpha$ such that $\lim_{\alpha\to 0}(T(\tilde{h}_\alpha) - I(\tilde{h}_\alpha)) = \lim_{\alpha\to 0}\sup_{\tilde{H}_\alpha}(T(\tilde{h}) - I(\tilde{h}))$. These reduced graphons converge to a reduced graphon $\tilde{h}_0 \in \tilde{H}_0$. Since T is continuous and I is lower semicontinuous,

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

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. □

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 semicontinuity of I, $\tilde{H}$ is a nonempty compact set. Theorem 1 gives an asymptotic formula for $\psi^e_{n,\alpha}$ but says nothing about the behavior of a typical random graph sampled from the constrained exponential model (11). In the unconstrained case (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 ψ 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 2.

Theorem 2. *Take $e \in [0,1]$. Let $\tilde{H}$ be defined as above. Let $\mathbb{P}^e_{n,\alpha}(G_n)$ (see (11)) be the conditional probability measure on $\mathcal{G}_n$. Then, for any $\eta > 0$ and sufficiently small α, there exist $C, \gamma > 0$ such that, for all large enough n,*

$$\mathbb{P}^e_{n,\alpha}(\delta_\square(\tilde{h}^{G_n}, \tilde{H}) \geq \eta) \leq C\mathrm{e}^{-n^2\gamma}.$$

Proof. We check that the conditional probability measure $\mathbb{P}^e_{n,\alpha}$ is well defined for all large enough n. It suffices to show that $\psi^e_{n,\alpha}$ is finite. But from (16), $\psi^e_{n,\alpha}$ 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 η-distance away from $\tilde{H}$,

$$\tilde{A}_\alpha = \{\tilde{h} \in \tilde{H}_\alpha\colon \delta_\square(\tilde{h}, \tilde{H}) \geq \eta\}. \tag{18}$$

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 semicontinuity of $T - I$, it follows that

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

From the proof of Theorem 1, we see that

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

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

Similarly, we have

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

This implies that, for sufficiently small α,

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

Choose $\varepsilon = \gamma$ and define $\tilde{H}_{\alpha,c}$ and R as in the proof of Theorem 1. Let $\tilde{A}_{\alpha,c} = \tilde{A}_\alpha \cap \tilde{H}_{\alpha,c}$. Then

$$\mathbb{P}^e_{n,\alpha}(\tilde{h}^{G_n} \in \tilde{A}_\alpha) \le \exp(-n^2\psi^e_{n,\alpha})|R| \sup_{c\in R}\exp(n^2(c+\gamma))|\tilde{A}^n_{\alpha,c}|.$$

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 1, from the above inequality, we have

$$\limsup_{n\to\infty}\frac{\log \mathbb{P}^e_{n,\alpha}(\tilde{h}^{G_n} \in \tilde{A}_\alpha)}{n^2} \le \sup_{c\in R}\Big(c + \gamma - \inf_{\tilde{h}\in\tilde{A}_{\alpha,c}} I(h)\Big) - \sup_{\tilde{h}\in\tilde{H}_{\alpha-\alpha^2}}(T(\tilde{h}) - I(\tilde{h})). \quad (19)$$

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

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

Substituting this in (19), we obtain

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

This completes the proof. □

4. An application

Theorems 1 and 2 in the previous section illustrate the importance of finding the maximizing graphons for $T - I$ subject to certain constraints. Similar optimization problems have also been studied in the context of upper tails of random graphs by Lubetzky and Zhao [16]. The optimizers 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 all possible constraints (referred to as macro analysis). However, as straightforward as it sounds, due to the myriad of structural possibilities of graphons, both the unconstrained (6) and constrained (13) optimization problems are not always explicitly solvable. So far major simplification has been achieved only in the 'attractive' case where the parameters $\beta_2, \dots, \beta_k$ are all nonnegative [8], [20], [26], and for k-star models [8], whereas a complete analysis of either (6) or (13) in the 'repulsive' region where the parameters $\beta_2, \dots, \beta_k$ are all negative has proved to be very difficult. In this section we will provide some phase transition results on the constrained repulsive edge-triangle exponential random graph model and discuss their possible generalizations. Using the same arguments, it is also possible to establish the phase transition in the attractive region of the parameter space. We make these notions precise in the following.

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

The unconstrained edge-triangle model is a two-parameter exponential random graph model obtained by taking H_1 to be a single edge and H_2 to be a triangle in (1). More explicitly, in the edge-triangle model, the probability measure $\mathbb{P}_n^\beta$ is

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

where $\beta = (\beta_1, \beta_2)$ are two real parameters, $e(G_n)$ and $t(G_n)$ are the edge and triangle densities of G_n, and ψ_n^β 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 ψ^e. From (13) we see that ψ^e depends on both parameters β_1 and β_2, however the β_1 dependence is linear: ψ^e is equal to $\beta_1 e$ plus a function independent of β_1. In particular, β_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 β_2.

To highlight this parameter dependence, in the following we will write ψ^e as ψ^{e,β_2} instead. We are particularly interested in the asymptotics of ψ^{e,β_2} when β_2 is negative, the so-called repulsive region. Naturally, varying β_2 allows one to adjust the influence of the triangle density of the graph on the probability distribution. The more negative the β_2, the more unlikely that graphs with a large number of triangles will be observed. When β_2 approaches $-\infty$, the most probable graph would likely be triangle free. At the other extreme, when β_2 is 0, 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 β_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 an 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 two-parameter models whose subgraph H_2 has chromatic number at least 3 has been proved by Aristoff and Radin [2], based on a symmetry breaking result from [8]. In Theorem 3 below we give a corresponding result in the constrained edge-triangle model. Its proof though is quite different from the parallel result in [2] and relies instead on some analysis arguments.

Theorem 3. *Consider the constrained repulsive edge-triangle exponential random graph model as described above. Let e be arbitrary but fixed. Let β_2 vary from 0 to $-\infty$. Then ψ^{e,β_2} is not analytic at for least one value of β_2.*

Proof. We first consider the case $e \le \frac{1}{2}$; the $e > \frac{1}{2}$ case is similar, see the comments at the end of the proof.

Let $e(\tilde{h}) \le \frac{1}{2}$ 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 (4). By (13), we have

$$\begin{aligned}\psi^{e,\beta_2} &= \sup_{\{\tilde{h}\in\tilde{\mathcal{W}}:\, e(\tilde{h})=e\}} (\beta_2 t(\tilde{h}) - I(\tilde{h})) \\ &= \sup_t \sup_{\{\tilde{h}\in\tilde{\mathcal{W}}:\, e(\tilde{h})=e, t(\tilde{h})=t\}} (\beta_2 t - I(\tilde{h})) \\ &= \sup_t (\beta_2 t + s(e,t)),\end{aligned} \tag{20}$$

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 (20) at the two extreme values

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

of β_2 first. Since I is convex, when $\beta_2 = 0$,

$$\psi^{e,0} = \sup_{\{\tilde{h}\in\tilde{\mathcal{W}}:\, e(\tilde{h})=e\}} (-I(\tilde{h})) \le -I(e)$$

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 $\beta_2 \le 0$, any maximizing graphon h for (20) will satisfy $t(\tilde{h}) \le e^3$. For the other extreme, take an arbitrary sequence $\beta_2^{(i)} \to -\infty$, 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 $\tilde{\mathcal{W}}$). We say that a graphon $h\colon [0,1]^2 \to [0,1]$ is symmetric bipodal if it is of the form

$$h(x,y) = \begin{cases} p & \text{if } x < \frac{1}{2} < y \text{ or } x > \frac{1}{2} > y, \\ q & \text{if } x, y < \frac{1}{2} \text{ or } x, y > \frac{1}{2}, \end{cases}$$

where p and q are constants taking values between 0 and 1. Suppose that $t(\tilde{h}) > 0$. Then by the continuity of t and the boundedness of I, $\lim_{i\to\infty} \psi^{e,\beta_2^{(i)}} = -\infty$. But this is impossible since $\psi^{e,\beta_2^{(i)}}$ is uniformly bounded below, as can be seen by considering the symmetric bipodal graphon h with $p = 2e$ and $q = 0$ as a test function, which corresponds to a complete bipartite graph with a $1 - 2e$ fraction of edges randomly deleted. 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 [18], [19]. From the convexity of I, Theorem 4.1 in [18] finds that, for $e \le \frac{1}{2}$, $s(e,0) = -I(2e)/2$ and this maximum is achieved only at the reduced symmetric bipodal graphon $\tilde{h}$ depicted above. Further utilizing properties of the Hermitian trace class operator, Theorem 1.1 in [19] states that, for any $e \in [0,1]$ and for $t \le e^3$,

$$s(e,e^3) - s(e,t) \ge c(e^3 - t)^{2/3} \quad \text{for some } c = c(e) > 0. \tag{21}$$

Thus, we have

$$\lim_{\beta_2\to-\infty} \psi^{e,\beta_2} = -\tfrac{1}{2} I(2e); \tag{22}$$

while (21) implies that, for $\beta_2 > -c(e)$ and $t < e^3$,

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

In other words, the constant graphon $h \equiv e$ still yields the maximum value for (20) for these small values of β_2. Thus regarded as a function of β_2, ψ^{e,β_2} is constant on the interval $(-c(e), 0)$ and $\psi^{e,\beta_2} = -I(e)$. This shows that ψ^{e,β_2} must lose its analyticity at least one β_2 as β_2 varies from 0 to $-\infty$, since otherwise we would have

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

in contradiction with (22).

For $e > \frac{1}{2}$, 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 [18], and their rate functions are strictly less than $-I(e)$, so the proof above goes through without change. □

The proof of Theorem 3 does not rely heavily on the definition of the edge-triangle model, except for the nondifferentiability of $s(e,t)$ at $t = e^3$ and the structure of the maximizing

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

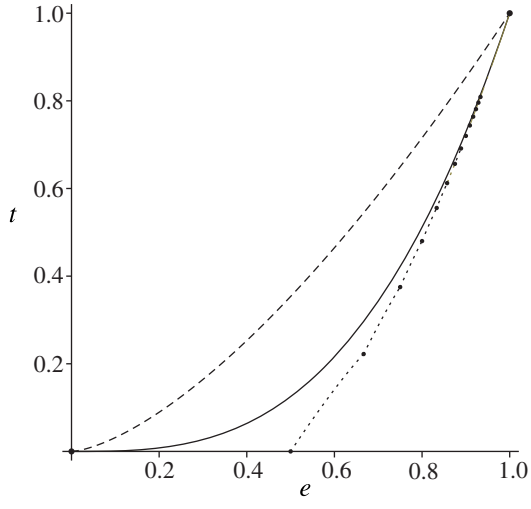


FIGURE 1: Region of attainable edge (e) and triangle (t) densities for graphons. The upper boundary is the curve $t = e^{3/2}$ (*dashed*) and the lower boundary is a piecewise algebraic curve (*dotted*) with infinitely many concave pieces; see [22]. The solid curve is the Erdős–Renyi curve $t = e^3$.

graphons at the two extreme values of β_2. The following extension of this theorem may not come as a surprise.

Theorem 4. *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 two-parameter exponential random graph model, where the probability measure $\mathbb{P}_n^{e,\beta_2}$ is given by*

$$\mathbb{P}_n^{e,\beta_2}(G_n) = \exp(n^2(\beta_2 t(H_2, G_n) - \psi_n^{e,\beta_2})). \tag{23}$$

Let the edge density e be fixed. Let the second parameter β_2 vary from 0 to $-\infty$. Then ψ^{e,β_2} loses its analyticity at least one value of β_2.

Proof. The proof of Theorem 3 carries over almost word for word when we incorporate the disordered Erdős–Rényi structure of the maximizing graphon at $\beta_2 = 0$, the nondifferentiability of $s(e, t)$ for a general H_2 [19], and the emergent multipartite structure of the maximizing graphon as $\beta_2 \to -\infty$ [8], [27]. □

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? In Theorem 5 we show what happens when the ideal edge density of the edge-triangle model is fixed at $\frac{1}{2}$ while the influence of the triangle densities is tuned through the parameter β_2.

Theorem 5. *Consider the constrained repulsive edge-triangle exponential random graph model as described at the beginning of Section 4. Fix $e = \frac{1}{2}$. Let β_2 vary from 0 to $-\infty$. Then $\psi^{1/2,\beta_2}$ is analytic everywhere except at a certain point β_2^c, where the derivative $\partial\psi^{1/2,\beta_2}/\partial\beta_2$ displays jump discontinuity.*

Proof. Setting $e = \frac{1}{2}$ in (20), we have

$$\psi^{1/2,\beta_2} = \sup_t\big(\beta_2 t + s\big(\tfrac{1}{2}, t\big)\big). \tag{24}$$

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

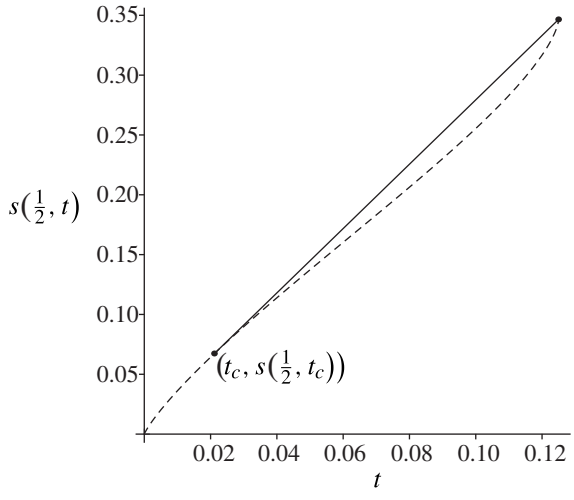


FIGURE 2: The graph of $s(\frac{1}{2}, t)$ below the Erdős–Rényi curve for the edge-triangle model (*dashed*). The convex hull of the region below the graph is delimited by the solid line segment and the portion of the graph to its left; this segment is the support line at the right endpoint ($t = \frac{1}{8},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

Since $\beta_2 \leq 0$, by the convexity of I, any maximizing graphon h for (24) must satisfy $t(\tilde{h}) \leq \frac{1}{8}$, that is, it must lie below the Erdős–Rényi curve $t = e^3$. Radin and Sadun [19] showed that on the line segment $e = \frac{1}{2}$ and $t \leq e^3$, the symmetric bipodal graphon

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

where $0 \leq \varepsilon = (\frac{1}{8} - t)^{1/3} \leq \frac{1}{2}$, maximizes $s(\frac{1}{2}, t)$, and that every maximizing graphon is of the form h_σ for some measure preserving bijection σ. Equivalently, the maximum value for (24) is achieved only at the reduced bipodal graphon $\tilde{h}$. See Figure 2 for the graph of $s(\frac{1}{2}, t)$.

Geometrically, the maximization problem in (24) involves finding the lowest half-plane with bounding line of slope $-\beta_2$ lying above the graph of $s(\frac{1}{2}, t)$. For $\beta_2 > \beta_2^c$ the boundary of this half-plane passes through the graph of $s(\frac{1}{2}, t)$ only at the right endpoint $(\frac{1}{8}, \frac{1}{2}\log 2)$. The critical value β_2^c is defined (as in Figure 2) as the first slope at which this half-plane intersects the curve at a different point. We let $(t_c, s(\frac{1}{2}, t_c))$ be this second point. At more negative values of β_2, the half-plane will hit the curve at points with t values below t_c.

In particular, this shows the nonanalyticity of $\psi^{1/2,\beta_2}$ as a function of β_2 at $\beta_2 = \beta_2^c$. The analyticity of $\psi^{1/2,\beta_2}$ elsewhere follows from concavity (and analyticity) of $s(\frac{1}{2}, t)$ below t_c. By Theorem 2, at $\beta_2 = \beta_2^c$, the maximizing reduced graphon $\tilde{h}$ for (24) transitions from being Erdős–Rényi with edge formation probability $\frac{1}{2}$ to symmetric bipodal with $\varepsilon_c = (\frac{1}{8} - t_c)^{1/3}$. The jump discontinuity in the derivative follows when we realize that $\partial\psi^{1/2,\beta_2}/\partial\beta_2 = t(\tilde{h})$, completing the proof. □

Numerical computations yield that β_2^c is approximately -2.7 and ε_c is approximately 0.47. By Theorem 2, this shows that as β_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 a first-order phase transition) is intrinsically tied to the *convexity* of $s(e, t)$ just below the Erdős–Rényi

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

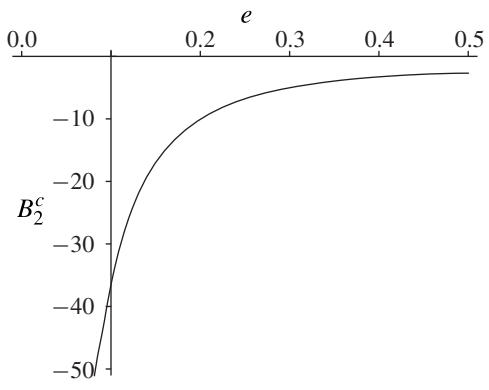


FIGURE 3: The (conjectural) graph of β_2^c as a function of e for the edge-triangle model, in the range $e \leq \frac{1}{2}$. The computation is based on the conjecture that the maximizing graphons in this region are symmetric and bipodal, see [21].

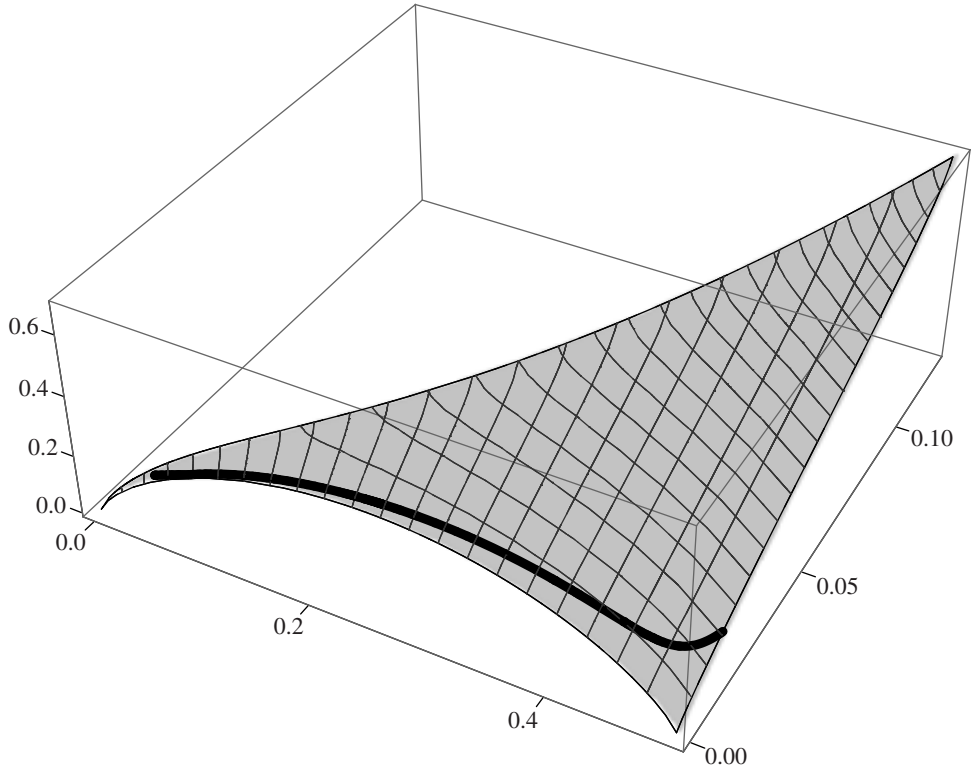


FIGURE 4: The (conjectural) graph of entropy $-I$ as a function of edge and triangle densities e, t in the region $e \leq \frac{1}{2}, t \leq e^3$. The critical curve (*bold*) defines t^c as a function of e. The computation is based on the conjecture that the maximizing graphons in this region are symmetric and bipodal, see [21].

curve $t = e^3$, thus we expect similar phase transition phenomena for general $e \neq \frac{1}{2}$ as well; see Figures 3 and 4. However, unlike in the $e = \frac{1}{2}$ case, where the symmetry of I (7) about $u = \frac{1}{2}$ contributes to a precise knowledge of the structure of the maximizing graphon, in general

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

cases, there is only empirical evidence concerning the structure of the maximizing graphons. See also [8] for related results in the unconstrained repulsive edge-triangle model.

5. Euler–Lagrange equations

We return to the constrained two-parameter family of exponential random graphs (23). 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 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 two-parameter model, they showed that h must be bounded away from 0 and 1 and, for almost all $(x, y) \in [0, 1]^2$,

$$h(x, y) = \frac{e^{2\sum_{i=1}^{2} \beta_i \Delta_{H_i} h(x,y)}}{1 + e^{2\sum_{i=1}^{2} \beta_i \Delta_{H_i} h(x,y)}}, \tag{25}$$

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

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

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

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

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 $\Delta_{H_2} h(x, y) = 3\int_0^1 h(x, z)h(y, z)\,\mathrm{d}z$. 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 6. *Consider the constrained two-parameter exponential random graph model (23). Let t_1 and t_2 be arbitrary but fixed homomorphism densities. Suppose that 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 β_1 and β_2 such that h satisfies (25) for almost all $(x, y) \in [0, 1]^2$.*

Proof. Graphons are bounded integrable functions on $[0, 1]^2$ so they are continuous outside a set of arbitrarily small measure. Let (x_i, y_i) for $i = 1, 2, 3$ be three points of $[0, 1]^2$. Inside a very small ball near (x_i, y_i), write $h = h_i + \bar{h}$, where h_i is the average of h in that ball. We infinitesimally perturb the values of h around (x_i, y_i), sending $h_i \to h_i + \mathrm{d}h_i$. Since $\bar{h}$ averages to 0 and is pointwise small, in computing t_1, t_2, and $-I$, terms involving $\bar{h}$ only

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

contribute to second order and may be ignored in the computation below. Then $(t_1, t_2, -I) \to (t_1, t_2, -I) + (\mathrm{d}t_1, \mathrm{d}t_2, -\mathrm{d}I)$, where

$$\begin{pmatrix} \mathrm{d}t_1 \\ \mathrm{d}t_2 \\ -\mathrm{d}I \end{pmatrix} = \begin{pmatrix} \Delta_{H_1} h_1 & \Delta_{H_1} h_2 & \Delta_{H_1} h_3 \\ \Delta_{H_2} h_1 & \Delta_{H_2} h_2 & \Delta_{H_2} h_3 \\ \frac{1}{2}\log(1/h_1 - 1) & \frac{1}{2}\log(1/h_2 - 1) & \frac{1}{2}\log(1/h_3 - 1) \end{pmatrix} \begin{pmatrix} \mathrm{d}h_1 \\ \mathrm{d}h_2 \\ \mathrm{d}h_3 \end{pmatrix}.$$

If the determinant of the above matrix is nonzero then there is a nontrivial deformation $(\mathrm{d}h_1, \mathrm{d}h_2, \mathrm{d}h_3)$ which increases $-I$ while leaving t_1 and t_2 fixed. So the maximizing graphon h must satisfy the condition that the determinant is 0. Recall that H_1 is a single edge and $\Delta_{H_1} h_i \equiv 1$. Without loss of generality, we assume that $h_1 \neq h_2$, since otherwise h is a constant graphon and our claim trivially follows. Thus, the first and third rows of the matrix are linearly independent and there must exist constants β_1 and β_2 such that

$$\Delta_{H_2} h_i = \beta_1 + \frac{\beta_2}{2} \log\left(\frac{1}{h_i} - 1\right). \tag{26}$$

Moreover, since β_1 and β_2 are determined by h_1 and h_2, we must have (26) for all points $(x_3, y_3) \in [0, 1]^2$. We recognize this requirement is equivalent to (25). □

Suppose that 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

$$\begin{pmatrix} \mathrm{d}t_1 \\ -\mathrm{d}I \end{pmatrix} = \begin{pmatrix} \Delta_{H_1} h_1 & \Delta_{H_1} h_2 \\ \frac{1}{2}\log(1/h_1 - 1) & \frac{1}{2}\log(1/h_2 - 1) \end{pmatrix} \begin{pmatrix} \mathrm{d}h_1 \\ \mathrm{d}h_2 \end{pmatrix}$$

Since the determinant is 0, 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 two) constraints $t(H_i, h) = t_i$ for $i = 1, \ldots, k$, in which case we would perturb the values of the graphon at $k+1$ points and form a $(k+1) \times (k+1)$ matrix.

Acknowledgements

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. They thank Charles Radin, Kui Ren, and Lorenzo Sadun for helpful conversations.

References

[1] Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11,** 581–598.
[2] Aristoff, D. and Radin, C. (2013). Emergent structures in large networks. *J. Appl. Prob.* **50,** 883–888.
[3] Bollobás, B. (2001). *Random Graphs*, 2nd edn. (Camb. Stud. Adv. Math. **73**). Cambridge University Press.
[4] Borgs, C., Chayes, J. T., Cohn, H. and Zhao, Y. (2014). An L^p theory of sparse graph convergence I. Limits, sparse random graph models, and power law distributions. Preprint. Available at https://arxiv.org/pdf/1401.2906./pdf.
[5] Borgs, C. *et al.* (2006). Counting graph homomorphisms. In *Topics in Discrete Mathematics* (Algorithms Combin. **26**). Springer, Berlin, pp. 315–371.
[6] Borgs, C. *et al.* (2008). Convergent sequences of dense graphs I: subgraph frequencies, metric properties and testing. *Adv. Math.* **219,** 1801–1851.
[7] Borgs, C. *et al.* (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. Math.* **176,** 151–219.

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press

[8] Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41,** 2428–2461.
[9] Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Prob.* **21,** 1400–1435.
[10] Chatterjee, S. and Varadhan, S. R. S. (2011). The large deviation principle for the Erdős-Rényi random graph. *Europ. J. Combinatronics* **32,** 1000–1017.
[11] Frank, O. and Strauss, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81,** 832–842.
[12] Häggström, O. and Jonasson, J. (1999). Phase transition in the random triangle model. *J. Appl. Prob.* **36,** 1101–1115.
[13] Hoover, D. (1982). Row-column exchangeability and a generalized model for probability. In *Exchangeability in Probability and Statistics,* eds G. Koch and F. Spizzichino. North-Holland, Amsterdam, pp. 281–291.
[14] Lovász, L. (2012). *Large Networks and Graph Limits.* American Mathematical Society, Providence.
[15] Lovász, L. and Szegedy B. (2006). Limits of dense graph sequences. *J. Combinatorial Theory B* **96,** 933–957.
[16] Lubetzky, E. and Zhao, Y. (2016). On the variational problem for upper tails in sparse random graphs. To appear in *Structures Algorithms.* Available at http://arxiv.org/abs/1402.6011.
[17] Newman, M. (2010). *Networks.* Oxford University Press.
[18] Radin, C. and Sadun, L. (2013). Phase transitions in a complex network. *J. Phys. A* **46,** 305002.
[19] Radin, C. and Sadun, L. (2013). Singularities in the entropy of asymptotically large simple graphs. *J. Statist. Phys.* **158,** 853–865.
[20] Radin, C. and Yin, M. (2013). Phase transitions in exponential random graphs. *Ann. Appl. Prob.* **23,** 2458–2471.
[21] Radin, C., Ren, K. and Sadun, L. (2014). The asymptotics of large constrained graphs. *J. Phys. A* **47,** 175001.
[22] Razborov, A. (2008). On the minimal density of triangles in graphs. *Combin. Prob. Comput.* **17,** 603–618.
[23] Touchette, H., Ellis, R. S. and Turkington, B. (2004). An introduction to the thermodynamic and macrostate levels of nonequivalent ensembles. *Physica A* **340,** 138–146.
[24] Van der Hofstad, R. (2014). *Random Graphs and Complex Networks.* Cambridge University Press.
[25] Wasserman, S. and Faust, K. (2010). *Social Network Analysis.* Cambridge University Press.
[26] Yin, M. (2013). Critical phenomena in exponential random graphs. *J. Statist. Phys.* **153,** 1008–1021.
[27] Yin, M., Rinaldo, A. and Fadnavis, S. (2013). Asymptotic quantization of exponential random graphs. *Ann. Appl. Prob.* **26,** 3251–3285.

https://doi.org/10.1017/jpr.2016.93 Published online by Cambridge University Press