TAME SPARSE EXPONENTIAL RANDOM GRAPHS

SUMAN CHAKRABORTY, REMCO VAN DER HOFSTAD, AND FRANK DEN HOLLANDER

Abstract. In this paper, we obtain a precise estimate of the probability that the sparse binomial random graph contains a large number of vertices in a triangle. The estimate of log of this probability is correct up to second order, and enables us to propose an exponential random graph model based on the number of vertices in a triangle. Specifically, by tuning a single parameter, we can with high probability induce any given fraction of vertices in a triangle. Moreover, in the proposed exponential random graph model we derive the large deviation principle for the number of edges. As a byproduct, we propose a consistent estimator of the tuning parameter.

1. Introduction

Background. Many real-world networks are sparse, while at the same time exhibiting transitivity (also called clustering), in the sense that two neighbours of the same vertex are more likely to also be neighbours of one another (see e.g., Newman [14], Rapoport [15], Serrano and Boguña [17], Watts and Strogatz [19]). As a result, many vertices in these graphs lie in triangles.

Exponential random graphs models (ERGM) are popular for modelling sparse real-world networks. Let $\mathcal{G}_n$ be the space of all simple graphs on the vertex set $[n]$, which has $2^{\binom{n}{2}}$ elements. An ERGM can be represented by its law

$$\mathbb{P}_T(G) = \frac{1}{Z_n(\beta)} \exp\left(\beta T(G)\right), \quad G \in \mathcal{G}_n, \tag{1.1}$$

where $T(G)$ is a real-valued function on the space of graphs, β is an appropriately chosen parameter (called the inverse temperature in statistical physics), and $Z_n(\beta)$ is the normalisation constant (called the partition function). In (1.1), the random variable $T(G)$ is a sufficient statistic, in the sense that G conditionally on $T(G) = t$ is *uniform* over all graphs G with $T(G) = t$. Examples of sufficient statistics T include linear combinations of subgraph counts, such as the number of edges, triangles, cycles, etc. ERGMs were first studied in Holland and Leinhardt [10], Frank and Strauss [7]. Several new sufficient statistics were introduced in Snijders, Pattison, Robins and Handcock [18].

The evaluation of the partition function $Z_n(\beta)$ is a fundamental (and often difficult) problem, and is closely related to an appropriate scaling of the inverse temperature β. In the dense regime, the first such result was obtained by Chatterjee and Diaconis [5]. While ERGMs are well-understood in the dense regime, there are hardly any results in the sparse regime (a recent result in the sparse regime appeared in Mukherjee [13], and a related model called the random triangle model was studied in Jonasson [12] and Häggström and Jonasson [8]). Unfortunately, even dense exponential random graphs are problematic, as shown in Bhamidi, Bresler and Sly [1]: either they locally look

Date: June 25, 2024.

1991 *Mathematics Subject Classification.* Primary: 05C80, 60F10, 05C30 62F12.

Key words and phrases. Random graphs, Nonlinear large deviations, Exponential random graphs, Consistent estimation.

like dense Erdős-Rényi random graphs, or the mixing times of Glauber dynamics or Metropolis-Hasting dynamics on them are exponentially large. Large-deviation-type estimates are the key to studying ERGMs. In dense regimes, this connection has been investigated in Chatterjee and Diaconis [5], Chatterjee and Dembo [4], Bhamidi, Chakraborty, Cranmer and Desmarais [2].

Goal and innovation of the present paper. In the present paper we continue our work in [3], where we investigated sparse ERGMs based on the number of vertices $V_T(G)$ that participate in triangles, and we computed the correct order of scaling of β, which turns out to be of the order $\log n$. Yet, we arrived at the disappointing conclusion that either there are very few vertices in triangles in the graph (when $\beta = a \log n$ with $a < \frac{1}{3}$), or virtually all vertices participate in triangles (when $\beta = a \log n$ with $a > \frac{1}{3}$). Both are unrealistic from a practical perspective. In this paper we consider the *critical case*, which corresponds to $\beta = \frac{1}{3} \log n + \theta$ for arbitrary $\theta \in \mathbb{R}$, and show that now $V_T(G)$ scales in a non-trivial manner, leading to sparse ERGMs with a tuneable fraction of vertices in triangles. We do this via a second-order non-linear large deviation analysis of $V_T(G)$ in the sparse Erdős–Rényi random graph, which is novel. We further prove a large deviation principle for the number of edges, showing that the model indeed is sparse. We propose several related ERGMs based on the number of vertices in triangles, and show how to consistently estimate the parameters in such models. We reiterate that there have been numerous studies to come up with sparse ERGMs [7, 9, 10, 11, 18] with a large number of triangles but as far as we know this is the first instance where an ERGM is rigorously shown to simultaneously satisfy the following three desirable properties: 1. a typical outcome is sparse, that is, the number of edges is linear in number of vertices with high probability, 2. the number of triangles in a typical outcome is linear number of triangles (and a linear number of vertices are part of a triangle), 3. consistent estimation of the parameter(s) is possible and the estimators are based on the sufficient statistic(s). Another advantage is that the estimators are explicit, thus the estimation is fast and one does not need to resort to simulations which is often time-consuming.

Organisation. This paper is organised as follows. In Section 2, we estimate the second-order of the large-deviation probabilities of the rare event that a sparse Erdős–Rényi random graph has a linear number of vertices in triangles, study the structure of the graph conditionally on this rare event, and provide proofs for our main results. In Section 3, we use these results, as well as the key insights developed in their proofs, to study exponential random graphs based on the number of vertices in triangles. We show that, for appropriate parameter choices, such models are *sparse*, i.e., lead to sparse exponential random graphs. In Section 4, we show how our main results can be used to *consistently* estimate the exponential random graph parameters. We close in Section 5 with a discussion and a list of open problems.

2. Large number of vertices in triangles

Let $G = (V(G), E(G))$ be a graph with vertex set $V(G)$ and edge set $E(G)$. Let $V_T(G)$ be the number of vertices that are part of a triangle in G. Throughout this paper, we write G_n for the Erdős–Rényi random graph $G(n, p_n)$, where $p_n = \lambda/n$. In this section, we study the probability that $V_T(G_n)$ is of order n, as well as the structure of the graph conditionally on this event.

Our first main result is an upper tail estimate for the sparse binomial random graph:

Theorem 2.1 (Large deviations for number of vertices in triangles)**.** *Let G_n be the Erdős–Rényi random graph with $p_n = \lambda/n$ and $\lambda \in (0, \infty)$. Then*

$$\mathbb{P}(V_T(G_n) \geq q) = \frac{n(n-1) \times \cdots \times (n-q+1)}{(3!)^{q/3}(q/3)!}\, p_n^q\, \mathrm{e}^{o(n^{19/20})}. \tag{2.1}$$

In particular, for $a \in (0,1)$,

$$\begin{aligned}&\log \mathbb{P}(V_T(G_n) \geq an)\\ &\quad = -n(1-a)\log(1-a) - (\tfrac{1}{3}an)\log(\tfrac{1}{3}an) - an(\tfrac{2}{3} + \tfrac{1}{3}\log 6) + an\log\lambda + o(n^{19/20}).\end{aligned} \tag{2.2}$$

Theorem 2.1 extends [3, Theorem 1.8] from first- to second-order large deviations. The second-order term will be important to settle a number of open problems, including the computation of the number of edges in the graph conditionally on the graph having many vertices in triangles, as well as to prove that most of the triangles are actually disjoint.

Our second main result says that if G_n is conditioned to have of the order n vertices in a triangle, then almost all the triangles are vertex disjoint. To state the result, we let $\mathrm{DT}_n(G_n)$ be the maximum number of vertex-disjoint triangles in the graph G_n:

Theorem 2.2 (Most triangles are disjoint conditionally on $V_T(G_n) = an$)**.** *For any* $\varepsilon > 0$,

$$\mathbb{P}\big(\mathrm{DT}_n(G_n) \leq \tfrac{1}{3}(a-\varepsilon)n \,|\, V_T(G_n) = an\big) \leq \exp\left(-\tfrac{1}{6}\varepsilon n\log n + Cn\right), \tag{2.3}$$

where C *is a constant that is independent of* ε.

Our third main result is the following concentration property on the number of edges in the graph:

Theorem 2.3 (Concentration of edges conditionally on $V_T(G_n) = an$)**.** *Conditionally on* $V_T(G_n) = an$*, the number of edges in* G_n *is concentrated around* $(a + \lambda/2)n$.

Organisation of the proof of Theorems 2.1–2.3. The proof of Theorems 2.1–2.3 is organised as follows. The upper bound of Theorem 2.1 is proved in Section 2.1, the lower bound in Section 2.2. In Section 2.3 we use the technique in the proof of Theorem 2.1 to show that, conditionally on the graph having many vertices in triangles, most of the triangles are disjoint, as stated in Theorem 2.2. In Section 2.4 the graph is shown to be sparse, as stated in Theorem 2.3.

2.1. **Upper bound in Theorem 2.1.** A central definition in the proof is that of a *q-basic graph* and its configuration:

Definition 2.4 (q-basic graphs)**.**

(a) A subgraph $G \subseteq K_n$ is called *q-basic* if $V_T(G) = q$ and $V_T(G\backslash e) < q$ for all edges $e \in G$. Here, $G\backslash e$ denotes the graph with the edge e removed.

(b) We can partition a q-basic graph (non-uniquely) into sets $(V_1, V_2, V_{31}, V_{32})$ in such a way that the following conditions are met:
   (1) V_1 consists of disjoint triangles, while $V\backslash V_1$ has no disjoint triangles;
   (2) V_2 consists of the ends of disjoint edges, where the ends of the edges are connected to a vertex in V_1, and $V\backslash(V_1 \cup V_2)$ has no such edges;
   (31) V_{31} consists of vertices that are connected to both ends of an edge between two triangles in V_1, or a so far unused edge between V_1 and V_2;
   (32) $V_{32} = V\backslash(V_1 \cup V_2 \cup V_{31})$.

(c) We say that a q-basic graph G has an $(\ell_1, \ell_2, \ell_{31}, \ell_{32})$ configuration if there is a graph partition $(V_1, V_2, V_{31}, V_{32})$ as described in part (b) with $|V_i| = \ell_i$ for $i = 1, 2, 31, 32$. (Note that $q = \ell_1 + \ell_2 + \ell_{31} + \ell_{32}$.) ◀

In words, V_{32} consists of vertices that are connected to: (i) both ends of an edge of a triangle in V_1; or (ii) an edge between a vertex in V_2 and a vertex in V_1 that is used to create the triangle containing the edge in V_2; or (iii) an edge between a vertex in V_2 and a vertex in V_1 that is used to construct V_{31}.

It is clear that the number of edges in a q-basic graph with an $(\ell_1, \ell_2, \ell_{31}, \ell_{32})$ configuration is equal to $\ell_1 + \frac{3}{2}\ell_2 + 3\ell_{31} + 2\ell_{32}$. The number of q-basic subgraphs of the complete graph K_n with an $(\ell_1, \ell_2, \ell_{31}, \ell_{32})$ configuration is at most

$$\frac{n(n-1)\times\cdots\times(n-q+1)}{(3!)^{\ell_1/3}\,(\ell_1/3)!\,(2!)^{\ell_2/2}\,(\ell_2/2)!\,\ell_3!}\times \ell_1^{\ell_2/2}\times\binom{\ell_1^2/2+\ell_1\ell_2}{\ell_{31}}\times\binom{\ell_1+\ell_2+\ell_{31}}{\ell_{32}}, \tag{2.4}$$

where $\ell_3 = \ell_{31} + \ell_{32}$. Indeed, the factor $n!/(n-q)!$ equals the number of ways we can choose q vertices out of n. However, permutations of the 3 vertices per triangle, as well as the collection of 3 vertices in the $\ell_1/3$ triangles, leaves the q-basic graph unchanged, so we have to divide by $(3!)^{\ell_1/3}\,(\ell_1/3)!$. Similarly, permutations of the pairs of vertices in the ℓ_2 edges in V_2, and the ℓ_2 edges, leave the q-basic graph unchanged, so we have to divide by $(2!)^{\ell_2/2}\,(\ell_2/2)!$. In the same vein, we have to divide by $\ell_3!$. This gives us the number of ways to partition the vertex set into $(V_1, V_2, V_{31}, V_{32})$. Per edge in V_2, we have ℓ_1 choices for the common neighbour of its endpoints, which gives rise to $\ell_1^{\ell_2/2}$ possibilities. Per vertex in V_{31}, we have to choose a pair of vertices in V_1, or in V_1 and V_2, which gives rise to $\binom{\ell_1^2/2+\ell_1\ell_2}{\ell_{31}}$ possibilities. Per vertex in V_{32}, we need to choose an edge in a triangle in V_1, of which there are ℓ_1, or an edge between a vertex in V_2 and a vertex in V_1 that is used to create the triangle containing the edge in V_2, of which there are ℓ_2, or an edge between V_2 and V_1 that is used to construct V_{31}, of which there are ℓ_{31}. This gives rise to $\binom{\ell_1+\ell+2+\ell_{31}}{\ell_{32}}$ possibilities.

We estimate $\frac{1}{2}\ell_1^2 + \ell_1\ell_2 \le q^2 \le n^2$ and $\ell_1 + \ell_2 + \ell_{31} \le q \le n$, to get that (2.4) is bounded above by

$$\frac{n(n-1)\times\cdots\times(n-q+1)}{(3!)^{\ell_1/3}\,(\ell_1/3)!\,(2!)^{\ell_2/2}\,(\ell_2/2)!\,\ell_3!}\times\exp\big((\ell_2/2)\log n + 2\ell_{31}\log n + \ell_{32}\log n\big). \tag{2.5}$$

We estimate

$$\begin{aligned}
\mathbb{P}(V_T(G_n)\ge k_n) &= \sum_{q\ge k_n}\mathbb{P}(V_T(G_n)=k_n)\\
&\le \sum_{q\ge k_n}\mathbb{P}\big(G_{n,p_n}\text{ contains a }q\text{-basic subgraph}\big)\\
&\le \sum_{q\ge k_n}\sum_{\ell_1+\ell_2+\ell_{31}+\ell_{32}=q}\mathbb{P}\big(G_{n,p_n}\text{ contains a }q\text{-basic }(\ell_1,\ell_2,\ell_{31},\ell_{32})\text{ subgraph}\big)\\
&\le \sum_{q\ge k_n}\sum_{\ell_1+\ell_2+\ell_{31}+\ell_{32}=q} p^{\ell_1+3\ell_2/2+3\ell_{31}+2\ell_{32}}\,\#\{q\text{-basic }(\ell_1,\ell_2,\ell_{31},\ell_{32})\text{ subgraph}\}\\
&= \sum_{q\ge k_n} p^q \sum_{\ell_1+\ell_2+\ell_{31}+\ell_{32}=q} p^{\ell_2/2+2\ell_{31}+\ell_{32}}\,\#\{q\text{-basic }(\ell_1,\ell_2,\ell_{31},\ell_{32})\text{ subgraphs}\}.
\end{aligned} \tag{2.6}$$

Substituting (2.5) into (2.6), we get

$$\mathbb{P}(V_T(G_n)\ge k_n) \le n^5 \max_{\ell_1+\ell_2+\ell_{31}+\ell_{32}\ge k_n}\lambda^{\ell_2/2+2\ell_{31}+\ell_{32}}\frac{n(n-1)\times\cdots\times(n-q+1)}{(3!)^{\ell_1/3}(\ell_1/3)!(2!)^{\ell_2/2}(\ell_2/2)!\ell_3!}. \tag{2.7}$$

We are left with maximising

$$\frac{\lambda^{\ell_2/2+2\ell_{31}+\ell_{32}}}{(3!)^{\ell_1/3}\,(\ell_1/3)!\,(2!)^{\ell_2/2}\,(\ell_2/2)!\,\ell_3!}, \tag{2.8}$$

subject to $\ell_1+\ell_2+\ell_{31}+\ell_{32}=q$. Using Stirling's approximation $n!=n\log n-n+O(\log n)$, we get that (2.8) equals

$$\exp\Big((\ell_2/2+2\ell_{31}+\ell_{32})\log|\lambda|-(\ell_1/3)\log(\ell_1/3)+(\ell_1/3)(1-\log 3!) \\ -(\ell_2/2)\log(\ell_2/2)+(\ell_2/2)(1-\log 2!)-\ell_3\log\ell_3+\ell_3+O(\log(\ell_1\vee\ell_2\vee\ell_3))\Big). \tag{2.9}$$

We can trivially upper bound (2.9) by

$$\exp\Big((\ell_2/2+2\ell_3)(\log\lambda)_+-(\ell_1/3)\log(\ell_1/3)+(\ell_1/3)(1-\log 3!) \\ -(\ell_2/2)\log(\ell_2/2)+(\ell_2/2)(1-\log 2!)-\ell_3\log\ell_3+\ell_3+O(\log(\ell_1\vee\ell_2\vee\ell_3))\Big), \tag{2.10}$$

where $x_+=\max\{x,0\}$.

In order to investigate this bound further, we state a technical lemma:

Lemma 2.5 (Variational problem)**.** *For large enough q, the minimum value of*

$$f(x_1,x_2,x_3)=\tfrac{1}{3}x_1\log x_1+\tfrac{1}{2}x_2\log x_2+x_3\log x_3+c_1x_1+c_2x_2+c_3x_3, \tag{2.11}$$

subject to $x_1+x_2+x_3=q$, is attained when $x_1\geq q-q^{9/10}$ and $x_2,x_3\leq q^{9/10}$. In particular, there exists a $C>0$ such that

$$\min_{x_1+x_2+x_3=q} f(x_1,x_2,x_3)\geq \tfrac{1}{3}q\log q+c_1q-Cq^{9/10}\log q. \tag{2.12}$$

Proof. We use the Lagrange multiplier method. Define

$$\Lambda=f(x_1,x_2,x_3)+\mu(q-x_1-x_2-x_3). \tag{2.13}$$

The partial derivatives of Λ with respect to x_1,x_2,x_3 are

$$\begin{aligned}\frac{\partial\Lambda}{\partial x_1}&=\tfrac{1}{3}\log x_1+\tfrac{1}{3}+c_1-\mu,\\ \frac{\partial\Lambda}{\partial x_2}&=\tfrac{1}{2}\log x_2+\tfrac{1}{2}+c_2-\mu,\\ \frac{\partial\Lambda}{\partial x_3}&=\log x_3+1+c_3-\mu.\end{aligned} \tag{2.14}$$

Setting these equal to zero, we get the following relation for the Lagrange multiplier μ:

$$\exp(3\mu-3c_1-1)\Big[1+\exp(-\mu+3c_1-2c_2)+\exp(-2\mu+3c_1-2c_3)\Big]=q. \tag{2.15}$$

The solution is bounded above by μ_U and bounded below by μ_L for large enough q, where μ_U and μ_L are solutions of the equations

$$\exp(3\mu_U-3c_1-1)=q,\qquad \exp(3\mu_L-3c_1-1)=q-q^{9/10}. \tag{2.16}$$

□

By Lemma 2.5, with $x_1=\ell_1, x_2=\ell_2, x_3=\ell_{31}+\ell_{32}, c_1=(1-\log 3!)/3$, for large enough q, (2.10) is bounded above by

$$\exp\Big(-\tfrac{1}{3}q\log(\tfrac{1}{3}q)+\tfrac{1}{3}(1-\log 3!)q+Cq^{9/10}\log q\Big). \tag{2.17}$$

This completes the proof of the upper bound in Theorem 2.1. □

2.2. **Lower bound in Theorem 2.1.** In this section, we prove the lower bound in Theorem 2.1. We abbreviate $s_n = an$, and assume that $3|an$. Then we have the following lower bound:

$$\mathbb{P}(V_T(G_{n,p_n}) \geq an) \geq \frac{n(n-1)\times\cdots\times(n-s_n+1)}{(3!)^{s_n/3}(s_n/3)!} \times \mathbb{P}(123, 456, \ldots, s_{n-2}s_{n-1}s_n \text{ form triangles and the rest of the graph is triangle free}). \tag{2.18}$$

Indeed, the prefactor counts the number of ways to choose s_n triples. The event that these triples form triangles and the event that the rest of the graph is triangle free are independent, and each has the same probability as when the triples are 123, 456, etc. Next, we rewrite

$$\begin{aligned}&\mathbb{P}(123, 456, \ldots, s_{n-2}s_{n-1}s_n \text{ form triangles and the rest of the graph is triangle free}) \\ &= p_n^{s_n}\mathbb{P}(\text{the rest of the graph is triangle free} \mid 123, 456, \ldots, s_{n-2}s_{n-1}s_n \text{ form triangles}).\end{aligned} \tag{2.19}$$

Define B_n to be the set of edges that are used to form the triangles $R_n = \{123, 456, \ldots, s_{n-2}s_{n-1}s_n\}$. Clearly, $|B_n| = s_n$. Conditionally on all the edges in B_n being present, the number of triangles in the graph G_n excluding R_n can be written as

$$W_n = \sum_{\{(i,j,k)\colon 1\leq i<j<k\leq n\}\setminus R_n} a_{ij}a_{jk}a_{ki}. \tag{2.20}$$

Hence

$$\begin{aligned}&\mathbb{P}(\text{the rest of the graph is triangle free} \mid 123, 456, \ldots, s_{n-2}s_{n-1}s_n \text{ form triangles}) \\ &= \mathbb{P}(W_n = 0) \geq \mathbb{P}(\mathrm{Poi}(\mathbb{E}[W_n]) = 0) - d_{\mathrm{TV}}(W_n, \mathrm{Poi}(\mathbb{E}[W_n])).\end{aligned} \tag{2.21}$$

We compute $\mathbb{E}[W_n]$ in two parts: one part in which a triangle contains an edge in one of the triangles, which contributes at most $s_n n p_n^2$ to the expectation, and the other part where the rest of the graph contributes at most $\frac{1}{6}n^3p_n^3$ to the expectation. This gives

$$\mathbb{E}[W_n] \leq s_n n p_n^2 + \tfrac{1}{6}n^3p_n^3 = a + \tfrac{1}{6}\lambda^3 + o(1), \tag{2.22}$$

since $p_n = \lambda/n$ and $s_n = an$, and hence

$$\mathbb{P}(\mathrm{Poi}(\mathbb{E}[W_n]) = 0) \geq \mathrm{e}^{-\mathbb{E}[W_n]} = \mathrm{e}^{-a-\lambda^3/6}(1 + o(1)). \tag{2.23}$$

To estimate the last term in (2.21), we rely on a coupling between sums of weakly-dependent Bernoulli variables and Poisson random variables. Consider a sum of Bernoulli random variables

$$W = \sum_{\alpha\in\mathcal{I}} X_\alpha, \qquad X_\alpha \sim \mathrm{Ber}(p_\alpha), \quad \alpha \in \mathcal{I}, \qquad w = \sum_{\alpha\in\mathcal{I}} p_\alpha, \tag{2.24}$$

where $\mathcal{I}$ is a set of indices. Let $\mathcal{N}_\alpha \subseteq \mathcal{I}$ be the set of indices such that X_α is independent of $\{X_\beta\colon \beta \notin \mathcal{N}_\alpha \cup \{\alpha\}\}$. We assume that $\alpha \notin \mathcal{N}_\alpha$ for all $\alpha \in \mathcal{I}$. The following theorem provides a bound on the total variation distance between W and $\mathrm{Poi}(w)$:

Theorem 2.6 (Chen-Stein coupling to Poisson)**.** *For the random variable W defined in (2.24),*

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(w)) \leq \min\{1, 1/w\} \sum_{\alpha\in\mathcal{I}} \left(p_\alpha^2 + \sum_{\beta\in\mathcal{N}_\alpha} \big(p_\alpha p_\beta + \mathbb{E}(X_\alpha X_\beta)\big)\right). \tag{2.25}$$

Proof. See [16, Theorem 4.7]. □

Now we can complete the proof of the lower bound in Theorem 2.1. We apply Theorem 2.6 with $\mathcal{I} = \{(i,j,k)\colon 1 \leq i < j < k \leq n\}\setminus R_n$. The terms in the sum in (2.20) are of two types:

- ▷ If one of the edges ij, jk, ki is from B_n, then that edge is present and therefore that term has a $\mathrm{Ber}(p_n^2)$ distribution.
- ▷ If none of the edges ij, jk, ki is from B_n, then that term has a $\mathrm{Ber}(p_n^3)$ distribution.

Note that two of the edges ij, jk, ki cannot be from B_n. Two triangles are dependent only when they share precisely one edge (they cannot share two edges). Therefore a term in W_n depends on at most $3n$ other terms. This gives that the cross expectation between two dependent terms is at most p_n^4 (which happens when the shared edge is from B_n). The bound in Theorem 2.6 is

$$\min\{1,1/w\}\sum_{\alpha\in\mathcal{I}} p_\alpha^2 + \min\{1,1/w\}\sum_{\alpha\in\mathcal{I}}\sum_{\beta\in\mathcal{N}_\alpha}\left(p_\alpha p_\beta + \mathbb{E}(X_\alpha X_\beta)\right). \tag{2.26}$$

In the case of W_n, $\mathcal{N}_\alpha$ is the set of triangles that share an edge with α that is not in B_n, since if this were not the case, then X_α and X_β would be independent. Furthermore, $\sum_{\alpha\in\mathcal{I}} p_i^2 \le n^3 p_n^4$. To estimate the second term, we distinguish between four cases, arising from α using an edge in B_n or not, and β being such that it shares an edge with α:

- ▷ **Case I:** If a term in W_n contains an edge from B_n, then the number of other triangles that share the same edge from B_n is n, and the cross expectation is p_n^4. Therefore this case contributes $s_n n^2(p_n^4 + p_n^4)$ to the second term.
- ▷ **Case II:** If a term in W_n contains an edge from B_n, then the number of other triangles that share some edge (that is not from B_n) is at most $2n$, and the cross expectation is p_n^4. Therefore this case contributes $s_n n^2(p_n^4 + p_n^4)$ to the second term as well.
- ▷ **Case III:** If a term in W_n does not contain an edge from B_n, then the number of other triangles that share some edge but does not contain an edge from B_n is at most $3n$, and the cross expectation is p_n^5. Therefore this case contributes at most $3n^4(p_n^6 + p_n^5)$ to the second term.
- ▷ **Case IV:** Consider a triangle that contains an edge from B_n. Then any other triangle that contains a different edge from B_n can depend on the first triangle only if both edges are from one triangle in R_n, and the cross expectation is p_n^3. Therefore this case contributes at most $s_n \times 2 \times n(p_n^4 + p_n^3)$ to the second term.

It follows that

$$d_{\mathrm{TV}}(W_n, \mathrm{Poi}(\mathbb{E}[W_n])) \le \min\{1, 1/\mathbb{E}[W_n]\}\, n^3 \left(p_n^4 + 3n\left(p_n^4 + p_n^4\right)\right). \tag{2.27}$$

Hence

$$d_{\mathrm{TV}}(W_n, \mathrm{Poi}(\mathbb{E}[W_n])) = o(1). \tag{2.28}$$

We conclude that the probability in (2.21) is bounded from below by

$$\mathrm{e}^{-a-\lambda^3/6}(1+o(1)) - d_{\mathrm{TV}}(W_n, \mathrm{Poi}(\mathbb{E}[W_n])) = \mathrm{e}^{-a-\lambda^3/6}(1+o(1)), \tag{2.29}$$

which remains strictly positive as $n\to\infty$. This proves that

$$\mathbb{P}(V_T(G_n) \ge an) \ge \frac{n(n-1)\times\cdots\times(n-s_n+1)}{(3!)^{s_n/3}(s_n/3)!} p_n^{s_n} c, \tag{2.30}$$

for some $c>0$, as required. □

2.3. **Most triangles are disjoint: proof of Theorem 2.2.** In this section, we use the technique developed for the proof of the upper bound in Theorem 2.1 to prove Theorem 2.2. For this we wish to evaluate

$$\mathbb{P}\big(\mathrm{DT}_n(G_n) \le \tfrac{1}{3}(a-\varepsilon)n \,\big|\, V_T(G_n) = an\big) = \frac{\mathbb{P}(\mathrm{DT}_n(G_n)) \le \frac{1}{3}(a-\varepsilon)n, V_T(G_n) \ge an)}{\mathbb{P}(V_T(G_n)\ge an)}. \tag{2.31}$$

Note that for the denominator we can use the estimate in Theorem 2.1. Obtaining an estimate for the numerator is more subtle, and for that we again use our earlier decomposition argument. The crucial observation is the following: the event $\{\mathrm{DT}_n(G_n) \le \frac{1}{3}(a-\varepsilon)n, V_T(G_n) \ge an\}$ implies that G_n admits an $(\ell_1, \ell_2, \ell_{31}, \ell_{32})$ decomposition with $\ell_1 \le (a-\varepsilon)n$. We obtain the following upper bound for any $\varepsilon > 0$:

$$\begin{aligned}
&\mathbb{P}\big(\mathrm{DT}_n(G_n) \le \tfrac{1}{2}(a-\varepsilon)n, V_T(G_n) \ge an\big) \\
&\quad = \sum_{q \ge an} \mathbb{P}\big(\mathrm{DT}_n(G_n) \le \tfrac{1}{3}(a-\varepsilon)n, V_T(G_n) = an\big) \\
&\le \sum_{q \ge an} \mathbb{P}\big(G_n \text{ contains a } q\text{-basic subgraph with } \ell_1 \le (a-\varepsilon)n\big) \\
&\quad \le \sum_{q \ge an} \sum_{\ell_1+\ell_2+\ell_{31}+\ell_{32}=q} \mathbb{P}\big(G_n \text{ contains a } q\text{-basic } \ell_1, \ell_2, \ell_{31}, \ell_{32} \text{ subgraph with } \ell_1 \le (a-\varepsilon)n\big).
\end{aligned} \tag{2.32}$$

This is bounded from above by

$$\begin{aligned}
&\sum_{q \ge an} \sum_{\ell_1+\ell_2+\ell_{31}+\ell_{32}=q} p^{\ell_1+3\ell_2/2+3\ell_{31}+2\ell_{32}} \#\{q\text{-basic } (\ell_1, \ell_2, \ell_{31}, \ell_{32}) \text{ subgraph with } \ell_1 \le (a-\varepsilon)n\} \\
&= \sum_{q \ge an} p^q \sum_{\ell_1+\ell_2+\ell_{31}+\ell_{32}=q} p^{\ell_2/2+2\ell_{31}+\ell_{32}} \#\{q\text{-basic } (\ell_1, \ell_2, \ell_{31}, \ell_{32}) \text{ subgraph with } \ell_1 \le (a-\varepsilon)n\}.
\end{aligned} \tag{2.33}$$

Next, we use (2.5) to obtain

$$\begin{aligned}
&\mathbb{P}\big(\mathrm{DT}_n(G_n) \le \tfrac{1}{3}(a-\varepsilon)n, V_T(G_n) \ge an\big) \\
&\le n^4 \sum_{q \ge an} p^q \max_{\ell_1+\ell_2+\ell_{31}+\ell_{32}=q, \ell_1 \le (a-\varepsilon)n} \lambda^{\ell_2/2+2\ell_{31}+\ell_{32}} \frac{n(n-1)\times\cdots\times(n-q+1)}{(3!)^{\ell_1/3}(\ell_1/3)!(2!)^{\ell_2/2}(\ell_2/2)!\ell_3!}.
\end{aligned} \tag{2.34}$$

Note that an identical calculation as for Lemma 2.5 yields that

$$\begin{aligned}
&\max_{\ell_1+\ell_2+\ell_{31}+\ell_{32}=q, \ell_1 \le \frac{1}{3}(a-\varepsilon)n} \lambda^{\ell_2/2+2\ell_{31}+\ell_{32}} \frac{n(n-1)\times\cdots\times(n-q+1)}{(3!)^{\ell_1/3}(\ell_1/3)!(2!)^{\ell_2/2}(\ell_2/2)!\ell_3!} \\
&\qquad \le \exp\left(-\tfrac{1}{3}(a-\varepsilon)n\log n - \tfrac{1}{2}(q-a+\varepsilon)n\log n + O(n)\right).
\end{aligned} \tag{2.35}$$

Therefore, inserting (2.35) into (2.34), we get

$$\begin{aligned}
&\mathbb{P}\big(\mathrm{DT}_n(G_n) \le \tfrac{1}{3}(a-\varepsilon)n, V_T(G_n) \ge an\big) \\
&\qquad \le \exp\left(-\tfrac{1}{3}(a-\varepsilon)n\log n - \tfrac{1}{2}(q-a+\varepsilon)n\log n + O(n)\right).
\end{aligned} \tag{2.36}$$

Finally, insert the estimates from Theorem 2.1 and (2.36) into (2.31), to get the claim. The argument even shows that $\mathbb{P}\big(\mathrm{DT}_n(G_n) \le \frac{1}{3}(a-\varepsilon)n, V_T(G_n) \ge an\big)$ decays like $\mathrm{e}^{-\delta n\log n}$ for some $\delta = \delta(a,\varepsilon) > 0$ for every $a, \varepsilon > 0$. □

2.4. **The graph is sparse: Proof of Theorem 2.3.** We will use the technique of generating functions to show that the graph is sparse. We need the following standard lemma:

Lemma 2.7 (Concentration in terms of moment generating functions)**.** *Let X_n be a random variable such that*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[\mathrm{e}^{tX_n}\right] = \phi(t), \qquad t \in \mathbb{R}, \tag{2.37}$$

for some $\phi\colon \mathbb{R} \to \mathbb{R}$ *that is twice differentiable at* 0 *and satisfies* $\phi(0) = 0$ *and* $\sup_{t\in[0,\delta]} |\phi''(t)| \le C_\delta < \infty$ *for some* $\delta > 0$. *Then* $\frac{X_n}{n} \to \phi'(0)$ *in probability as* $n \to \infty$.

Proof. We use the Markov inequality. Fix any $\varepsilon > 0$. Then, for all $t > 0$,

$$\begin{aligned}\mathbb{P}\left(\frac{X_n}{n} \ge \phi'(0) + \varepsilon\right) &= \mathbb{P}\left(t\frac{X_n}{n} \ge t\phi'(0) + t\varepsilon\right)\\ &= \mathbb{P}\left(\exp(tX_n) \ge \exp(tn\phi'(0) + t\varepsilon)\right)\\ &\le \exp(-tn\phi'(0) - tn\varepsilon)\,\mathbb{E}\left[\mathrm{e}^{tX_n}\right].\end{aligned} \tag{2.38}$$

By (2.37),

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left(\frac{X_n}{n} \ge \phi'(0) + \varepsilon\right) \le -t\phi'(0) - t\varepsilon + \phi(t). \tag{2.39}$$

Expanding ϕ, we get, for $t \in [0, \delta]$

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left(\frac{X_n}{n} \ge \phi'(0) + \varepsilon\right) \le -t\varepsilon + C_\delta t^2. \tag{2.40}$$

Now we can choose t small enough to complete the proof of the upper bound. The lower bound works in the same way. □

The following lemma gives us control over the number of edges in G_n given $V_T(G_n)$:

Lemma 2.8 (Conditional edge moment generating function)**.** *For every* $a \in (0, 1]$ *and* $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{E}\big[\mathrm{e}^{te(G_n)} \mid V_T(G_n) = an\big] = \tfrac{1}{2}\lambda(\mathrm{e}^t - 1) + at. \tag{2.41}$$

Proof. For fixed $t \in \mathbb{R}$, compute

$$\mathbb{E}\left[\mathrm{e}^{te(G_n)}\mathbb{1}_{\{V_T(G_n)=an\}}\right] = \sum_{G\colon V_T(G)=an} \mathrm{e}^{te(G)}\, p_n^{e(G)}(1-p_n)^{\binom{n}{2}-e(G)}. \tag{2.42}$$

We can rewrite this relation as

$$\mathbb{E}\big[\mathrm{e}^{te(G_n)}\mathbb{1}_{\{V_T(G_n)=an\}}\big] = (\phi_n(t))^{\binom{n}{2}} \sum_{G\colon V_T(G)=an} p_n(t)^{e(G)}(1-p_n(t))^{\binom{n}{2}-e(G)}, \tag{2.43}$$

where

$$\phi_n(t) = 1 - p_n + \mathrm{e}^t p_n, \qquad p_n(t) = \frac{\mathrm{e}^t p_n}{1 - p_n + \mathrm{e}^t p_n}. \tag{2.44}$$

Hence

$$\mathbb{E}\big[\mathrm{e}^{te(G_n)}\mathbb{1}_{\{V_T(G_n)=an\}}\big] = (\phi_n(t))^{\binom{n}{2}}\, \mathbb{P}_{p_n(t)}(V_T(G_n) = an). \tag{2.45}$$

Using (2.45), we get

$$\mathbb{E}\big[\mathrm{e}^{te(G_n)} \mid V_T(G_n) = an\big] = \frac{\mathbb{E}\left[\exp(te(G_n))\mathbb{1}_{\{V_T(G_n)=an\}}\right]}{\mathbb{P}(V_T(G_n) = an)} = (\phi_n(t))^{\binom{n}{2}}\, \frac{\mathbb{P}_{p_n(t)}(V_T(G_n) = an)}{\mathbb{P}(V_T(G_n) = an)}. \tag{2.46}$$

At this point we use the estimate from (2.17), to get

$$\left(\frac{p_n(t)}{p_n}\right)^{an-(an)^c} \le \frac{\mathbb{P}_{p_n(t)}(V_T(G_n) = an)}{\mathbb{P}(V_T(G_n) = an)} \le \left(\frac{p_n(t)}{p_n}\right)^{an+(an)^{9/10}}. \tag{2.47}$$

Combining (2.46)–(2.47), we get the claim. □

Proof of Theorem 2.3. We use Lemma 2.7 with $X_n = e(G_n)$ conditionally on $V_T(G_n) = an$. In this case $\phi(t) = \frac{1}{2}\lambda(\mathrm{e}^t - 1) + at$, which satisfies the assumptions in Lemma 2.7. □

From Lemma 2.7, we can also conclude that the number of edges, conditionally on $V_T(G_n) = an$, satisfied a large deviation principle:

Corollary 2.9 (LDP for the number of edges)**.** *Conditionally on $V_T(G_n) = an$, the number of edges $e(G_n)$ satisfies a large deviation principle with rate n and with good rate function*

$$I_\lambda(x) = (x-a)\log\left(\frac{x-a}{\lambda/2}\right) - (x-a) + 1. \tag{2.48}$$

The rate function $x \mapsto I_\lambda(x)$ is the large deviation rate function of $na + e(G_n)$, where now G_n is an *unconditional* Erdős–Rényi random graph. This suggests that the conditioning on $V_T(G_n) = an$ simply adds $\frac{1}{3}an$ vertex-disjoint triangles, and the remainder of the graph still has that every edge is present independently with probability $p_n = \lambda/n$.

Proof. The claim follows directly from the Gärtner–Ellis Theorem (see [6, Theorem 2.3.6]). □

3. Exponential random graph models

In this section we consider exponential random graph models based on the number of vertices in triangles. In Section 3.1, we discuss our main results, which are of two types. In the first, we bias by $n \log n$ times the number of vertices in triangles (a setting which we call "linear tilting"). In the second, we bias by $n \log n$ times a function of the number of vertices in triangles (a setting which we call "functional tilting"). We investigate the asymptotics of the partition function, as well as the number of vertices in triangles and the number of edges, arising in the exponential random graph. In Section 3.2, we prove our main results for the linear tilting; in Section 3.3 for the functional tilting.

3.1. Results for exponential random graphs.

Linear tilting. Consider the exponential random graph model $\mathbb{P}_\beta$ defined by

$$\frac{d\mathbb{P}_\beta}{d\mathbb{P}} = \frac{1}{Z_n(\beta)}\,\mathrm{e}^{\beta \log n V_T(G_n)}, \tag{3.1}$$

where $\mathbb{P}$ is the measure of the Erdős–Rényi random graph with $p_n = \lambda/n$. This model was investigated in [3], where it was proved that $V_T(G_n)/n \xrightarrow{\mathbb{P}_\beta} 0$ for $\beta < \frac{1}{3}$ and $V_T(G_n)/n \xrightarrow{\mathbb{P}_\beta} 1$ for $\beta > \frac{1}{3}$. We investigate the critical window of this phase transition, for which we take

$$\beta = \frac{1}{3} + \frac{\theta}{\log n}. \tag{3.2}$$

It turns out that for this choice of parameter, the number of vertices in triangles concentrates on a non-trivial value $a^\star \in (0,1)$ that depends on θ:

Theorem 3.1 (Vertex-in-triangles exponential random graphs)**.** *Consider the exponential random graph in (3.1), with $\beta = \frac{1}{3} + \frac{\theta}{\log n}$. Then, for all $\theta \in \mathbb{R}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log Z_n(\beta) = \max_{a\in[0,1]} \Lambda(a,\theta,\lambda), \tag{3.3}$$

with

$$\Lambda(a,\theta,\lambda) = \theta a - (1-a)\log(1-a) - (\tfrac{1}{3}a)\log(\tfrac{1}{3}a) - \tfrac{2}{3}a - \tfrac{1}{3}a\log 6 + a\log\lambda. \tag{3.4}$$

Furthermore,

$$\frac{V_T(G_n)}{n} \xrightarrow{\mathbb{P}_\beta} a^\star, \tag{3.5}$$

where $a^\star$ is the unique maximizer of the variational problem in (3.3), and

$$\frac{E(G_n)}{n} \xrightarrow{\mathbb{P}_\beta} \frac{\lambda}{2} + a^\star. \tag{3.6}$$

Functional tilting. In this section, we extend Theorem 3.1 to exponential random graphs based on the number of vertices in triangles. Pick a function $g\colon [0,1] \to \mathbb{R}$. Consider the exponential random graph model $\mathbb{P}_g$ defined by

$$\frac{d\mathbb{P}_g}{d\mathbb{P}} = \frac{1}{Z_n(g)} \, \mathrm{e}^{n \log n g(V_T(G_n)/n)}, \tag{3.7}$$

where $\mathbb{P}$ is as before. We show that, under some mild assumptions on g, we can asymptotically evaluate the normalizing constant $Z_n(g)$, as well as the number of vertices in triangles and the number of edges:

Theorem 3.2 (Functional exponential random graphs)**.** *Let $g\colon [0,1] \to \mathbb{R}$ be a function such that all maxima of $a \mapsto g(a) - \frac{1}{3}a$ are attained in $(0,1)$. Let $a^\star \in (0,1)$ be any of the maximizers. Then*

$$\lim_{n\to\infty} \frac{1}{n \log n} \log Z_n(g) = g(a^\star) - \tfrac{1}{3}a^\star. \tag{3.8}$$

Further, if the maximum of $g(a) - \frac{1}{3}a$ is uniquely *attained at some $a^\star \in (0,1)$, then*

$$\frac{V_T(G_n)}{n} \xrightarrow{\mathbb{P}_g} a^\star, \tag{3.9}$$

while

$$\frac{E(G_n)}{n} \xrightarrow{\mathbb{P}_g} a^\star + \frac{\lambda}{2}. \tag{3.10}$$

Here is a concrete example of a function g for which Theorem 3.2 applies.

Corollary 3.3 (Example of functional tilting)**.** *Let $g(x) = \beta x^\alpha$ for some $\alpha \in (0,1)$, $\beta > 0$, and $3\alpha\beta > 1$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log Z_n(g) = \beta(3\alpha\beta)^{-\frac{\alpha}{\alpha-1}} - \tfrac{1}{3}(3\alpha\beta)^{-\frac{1}{\alpha-1}}. \tag{3.11}$$

Proof. Let $G(x) = g(x) - \frac{1}{3}x$. Note that $G'(x) = \beta\alpha x^{\alpha-1} - \frac{1}{3}$ and $G''(x) = \beta\alpha(\alpha-1)x^{\alpha-2}$. Therefore G attains its maximum in $(0,1)$ (since $3\alpha\beta > 1$). Therefore the claim follows from Theorem 3.2. □

Remark 3.4 (LDP for the number of edges)**.** In Theorems 3.1 and 3.2, we can extend the asymptotics of the number of edges to a large deviation principle, as in Corollary 2.9. The proof of this extension follows directly from the Gärtner–Ellis Theorem (see [6, Theorem 2.3.6]), and is therefore identical to the proof of Corollary 2.9. ◀

In the remainder of the section, we give the proofs of Theorems 3.1 and 3.2. The proofs are organised in the same way. First, we investigate the asymptotics of the partition function in terms of a variational problem. Afterwards, we conclude that the number of vertices in triangles divided by n converges to the maximizer of this variational problem. Finally, we compute the moment generating function for the number of edges to conclude the proof.

3.2. **Linear tilting.** The asymptotics for partition function is identified in the following lemma:

Lemma 3.5 (Partition function for the linear tilting)**.** *For every* $\theta \in \mathbb{R}$,

$$\log Z_n\big(\tfrac{1}{3}+\tfrac{\theta}{\log n}\big) = n \max_{0\le a\le 1} \Lambda(a,\theta,\lambda) + o(n^{19/20}). \tag{3.12}$$

Proof. By Theorem 2.1,

$$\begin{aligned}&\log \mathbb{P}(V_T(G_n)=an)\\ &= -n(1-a)\log(1-a) - \tfrac{1}{3}an\log n - \tfrac{1}{3}an\log(\tfrac{1}{3}a) - \tfrac{2}{3}an - \tfrac{1}{3}an\log 6 + an\log\lambda + o(n^{19/20}).\end{aligned} \tag{3.13}$$

We write

$$\mathbb{E}\left(\mathrm{e}^{\beta\log n V_T(G_n)}\right) = \sum_{q=0}^{n} \mathrm{e}^{\beta\log nq}\,\mathbb{P}\left(V_T(G_n)=q\right).$$

For $p_n=\lambda/n$, using the estimate in (3.13), we obtain

$$\begin{aligned}\mathbb{E}\left(\mathrm{e}^{\beta\log n V_T(G_n)}\right) &= n \sup_{0\le a\le 1}\\ &\times \exp\Big(\beta an\log n - n(1-a)\log(1-a) - \tfrac{1}{3}an\log(\tfrac{1}{3}an) - an(\tfrac{2}{3}+\tfrac{1}{3}\log 6 - \log\lambda) + o(n^{19/20})\Big).\end{aligned} \tag{3.14}$$

Substitute $\beta = \frac{1}{3}+\frac{\theta}{\log n}$, to get

$$\mathbb{E}\left(\mathrm{e}^{\beta\log n V_T(G_n)}\right) = n\exp\left(n\max_{0\le a\le 1}\Lambda(a,\theta,\lambda) + o(n^{19/20})\right). \tag{3.15}$$

□

Proof of Theorem 3.1. Fix an $\varepsilon>0$. We wish to upper bound $\mathbb{P}_\beta(V_T(G_n)\notin[(a^\star-\varepsilon)n,(a^\star+\varepsilon)n])$. This can be written as

$$\mathbb{P}_\beta\left(V_T(G_n)\notin[(a^\star-\varepsilon)n,(a^\star+\varepsilon)n]\right) = \frac{1}{Z_n(\beta)}\sum_{q\notin[(a^\star-\varepsilon)n,(a^\star+\varepsilon)n]} \mathrm{e}^{\beta\log nq}\,\mathbb{P}(V_T(G_n)=q). \tag{3.16}$$

Define

$$\delta = \Lambda(a^\star,\theta,\lambda) - \max_{a\notin[(a^\star-\varepsilon),(a^\star+\varepsilon)]}\Lambda(a,\theta,\lambda). \tag{3.17}$$

It is easy to check that $\delta>0$, since $\Lambda(\cdot,\theta,\lambda)$ is strictly concave and therefore has a unique maximum. Now we set $\beta=\frac{1}{3}+\frac{\theta}{\log n}$ and use (3.13), to obtain that $\log\sum_{q\notin[(a^\star-\varepsilon)n,(a^\star+\varepsilon)n]}\mathrm{e}^{\beta\log nq}\,\mathbb{P}(V_T(G_n)=q)$ is bounded above by $n(\frac{1}{4}\delta+\max_{a\notin[(a^\star-\varepsilon),(a^\star+\varepsilon)]}\Lambda(a,\theta,\lambda))$ for large enough n. Then use (3.12) to get $\log Z_n\big(\frac{1}{3}+\frac{\theta}{\log n}\big)\ge n\left(-\frac{1}{4}\delta+\Lambda(a^\star,\theta,\lambda)\right)$. Combining these estimates, we get

$$\log\mathbb{P}_{(\frac{1}{3}+\frac{\theta}{\log n})}\left(V_T(G_n)\notin[(a^\star-\varepsilon)n,(a^\star+\varepsilon)n]\right) \le -\tfrac{1}{2}n\delta. \tag{3.18}$$

Let us next compute the following moment generating function for the number of edges under the measure $\mathbb{P}_\beta$ defined in (3.1). For $t\in\mathbb{R}$,

$$\mathbb{E}_\beta\left[\mathrm{e}^{te(G_n)}\right] = \frac{1}{Z_n(\beta)}\mathbb{E}\left(\mathrm{e}^{te(G_n)+\beta\log nV_T(G_n)}\right) = \frac{(\phi_n(t))^{\binom{n}{2}}}{Z_n(\beta)}\mathbb{E}_{p_n(t)}\left(\mathrm{e}^{\beta\log nV_T(G_n))}\right), \tag{3.19}$$

where $\mathbb{E}_{p_n(t)}$ is the expectation w.r.t. the Erdős–Rényi random graph with $p=p_n(t)$.

Now using (3.12), we get

$$\begin{aligned}&(\phi_n(t))^{\binom{n}{2}} \exp\left(n \max_{0\le a\le 1} \Lambda(a,\theta,np_n(t)) - n \max_{0\le a\le 1} \Lambda(a,\theta,\lambda) + o(n^{19/20})\right) \le \mathbb{E}_\beta\left[\mathrm{e}^{te(G_n)}\right]\\ &\le (\phi_n(t))^{\binom{n}{2}} \exp\left(n \max_{0\le a\le 1} \Lambda(a,\theta,np_n(t)) - n \max_{0\le a\le 1} \Lambda(a,\theta,\lambda) + o(n^{19/20})\right).\end{aligned} \tag{3.20}$$

Finally, simplifying (3.20), we obtain

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}_\beta\left[\mathrm{e}^{te(G_n)}\right] = \tfrac{1}{2}\lambda(\mathrm{e}^t - 1) + \max_{0\le a\le 1} \Lambda(a,\theta+t,\lambda) - \max_{0\le a\le 1} \Lambda(a,\theta,\lambda).$$

Differentiating the right-hand side of the last display with respect to t at $t=0$, we find $\lambda/2 + a^\star$. (The first term is easy while the second term is handled in Lemma 3.6 below.) Therefore $\frac{E(G_n)}{n} \xrightarrow{\mathbb{P}_\beta} \frac{\lambda}{2} + a^\star$ by Lemma 2.7. □

Lemma 3.6. *Let* $\Lambda_{\max}(t) = \max_{0\le a\le 1} \Lambda(a,\theta+t,\lambda)$. *Then*

$$\frac{\partial \Lambda_{\max}(t)}{\partial t}\Big|_{t=0} = a^\star.$$

Proof. Recall that

$$\Lambda(a,\theta,\lambda) = \theta a - (1-a)\log(1-a) - (\tfrac{1}{3}a)\log(\tfrac{1}{3}a) - \tfrac{2}{3}a - \tfrac{1}{3}a\log 6 + a\log\lambda.$$

It is easy to see that Λ is concave in a and thus has a unique maximum. Define $a^\star(t) = \arg\max_{0\le\le 1} \Lambda(a,\theta+t,\lambda)$. Also note that $a^\star(t)$ must be the unique solution of the equation $\frac{\partial\Lambda}{\partial a} = 0$. More precisely,

$$\theta + t + \log(1-a^\star(t)) - \tfrac{1}{3}\log(\tfrac{1}{3}a^\star(t)) - \tfrac{1}{3}\log 6 + \log\lambda = 0. \tag{3.21}$$

Clearly $a^\star(0) = a^\star$, and

$$\Lambda_{\max}(t) = \max_{0\le a\le 1} \Lambda(a,\theta+t,\lambda) = \Lambda(a^\star(t),\theta+t,\lambda). \tag{3.22}$$

We can implicitly differentiate $a^\star(t)$ with respect to t and use (3.21), to obtain

$$\frac{\partial a^\star(t)}{\partial t} = \frac{3a^\star(t)(1-a^\star)}{1+2a^\star}. \tag{3.23}$$

We are now ready to evaluate $\frac{\partial\Lambda_{\max}(t)}{\partial t}$. Using (3.22), we get

$$\begin{aligned}\frac{\partial\Lambda_{\max}(t)}{\partial t} &= \frac{\partial\Lambda(a,\theta+t,\lambda)}{\partial a}\Big|_{a=a^\star(t)} \times \frac{\partial a^\star(t)}{\partial t} + \frac{\partial\Lambda(a,\theta+t,\lambda)}{\partial\theta}\Big|_{a=a^\star(t)}\\ &= \Big(\theta + t + \log(1-a^\star(t)) - \tfrac{1}{3}\log(\tfrac{1}{3}a^\star(t)) - \tfrac{1}{3}\log 6 + \log\lambda\Big)\frac{3a^\star(t)(1-a^\star(t))}{1+2a^\star(t)} + a^\star(t).\end{aligned} \tag{3.24}$$

Therefore (3.24) yields $\frac{\partial\Lambda_{\max}(t)}{\partial t}\big|_{t=0} = a^\star(0) = a^\star$. □

3.3. **Functional tilting.** The following lemma identifies the asymptotics of the partition sum:

Lemma 3.7 (Partition function for the functional tilting)**.** *Let* $g\colon [0,1] \to \mathbb{R}$ *be a function such that all maxima of* $a \mapsto g(a) - \frac{1}{3}a$ *are attained in* $(0,1)$. *Let* $a^\star \in (0,1)$ *be any of the maximizers. Then*

$$\lim_{n\to\infty} \frac{1}{n\log n} \log Z_n(g) = g(a^\star) - \tfrac{1}{3}a^\star, \tag{3.25}$$

and $V_T(G_n)/n \xrightarrow{\mathbb{P}_g} a^\star$.

Proof. Using (2.17), we get

$$\begin{aligned}\mathbb{E}\left(\mathrm{e}^{n\log n\, g(\frac{V_T(G_n)}{n})}\right) &= \sum_{q=0}^{n} \mathrm{e}^{n\log n g(\frac{q}{n})} \exp\left(-\tfrac{1}{3}q\log n + O(n)\right)\\ &\leq n\exp\left(n\log n \sup_{a\in[0,1]}(g(a)-\tfrac{1}{3}a) + O(n)\right).\end{aligned} \tag{3.26}$$

By assumption, the supremum is attained at $a^\star \in (0,1)$. Therefore we obtain the following lower bound as well:

$$\mathbb{E}\left(\mathrm{e}^{n\log n g(\frac{V_T(G_n)}{n})}\right) \geq \exp\left(n\log n\,(g(a^\star)-\tfrac{1}{3}a^\star) + O(n)\right). \tag{3.27}$$

Also, the main contribution comes from $V_T(G_n) \in [(a^\star-\varepsilon)n, (a^\star+\varepsilon)n]$. Combining (3.26)–(3.27), we get the claim. □

The limit of the number of edges is slightly more involved. The next lemma identifies its moment generating function, as in Lemma 2.8:

Lemma 3.8 (Edge moment generating function for functional tilting)**.** *Let $g\colon [0,1]\to\mathbb{R}$ be any function such that $a\mapsto g(a)-\frac{1}{3}a$ is uniquely maximized at $a^\star\in(0,1)$. Let $\mathbb{E}_g$ be the expectation with respect to $\mathbb{P}_g$ defined in (3.7). Then, for any $t\in\mathbb{R}$,*

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}_g\left[\mathrm{e}^{te(G_n)}\right] = \tfrac{1}{2}\lambda(\mathrm{e}^t-1)+a^\star t. \tag{3.28}$$

Proof. We wish to compute

$$\mathbb{E}_g\left[\mathrm{e}^{te(G_n)}\right] = \frac{1}{Z_n(g)}\,\mathbb{E}\left(\exp\left(te(G_n)+n\log n\, g(\tfrac{V_T(G_n)}{n})\right)\right). \tag{3.29}$$

This relation can be rewritten as

$$\mathbb{E}_g\left[\mathrm{e}^{te(G_n)}\right] = (\phi_n(t))^{\binom{n}{2}}\,\frac{1}{Z_n(g)}\,\mathbb{E}_{p_n(t)}\left(\exp\left(n\log n\, g(\tfrac{V_T(G_n)}{n})\right)\right). \tag{3.30}$$

Write

$$\mathbb{E}_{p_n(t)}\left(\exp\left(n\log n\, g(\tfrac{V_T(G_n)}{n})\right)\right) = \sum_{q=0}^{n}\mathrm{e}^{n\log n g(\frac{q}{n})}\,\mathbb{P}_{p_n(t)}(V_T(G_n)=q). \tag{3.31}$$

Using the estimate in (2.2), we obtain that the sum is bounded above by

$$\begin{aligned}n\sup_{0\leq a\leq 1}\exp\Big(&n\log n\,(g(a)-\tfrac{1}{3}a) - n(1-a)\log(1-a) - (\tfrac{1}{3}an)\log(\tfrac{1}{3}a)\\ &-\tfrac{2}{3}an - \tfrac{1}{3}an\log 6 + an\log(\mathrm{e}^t\lambda) + o(n^{19/20})\Big).\end{aligned} \tag{3.32}$$

In order to study this function, we need the following technical lemma:

Lemma 3.9. *Let $f_1\colon [0,1]\to\mathbb{R}$ and $f_2\colon [0,1]\to\mathbb{R}$ be two functions such that f_1 uniquely attains its maximum at $a^\star\in[0,1]$ and f_1, f_2 are continuous on $[0,1]$. Let $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ be sequences of non-negative real numbers such that $b_n/a_n\to 0$ as $n\to\infty$. Then, for every fixed $\varepsilon>0$, $a_nf_1(a)+b_nf_2(a)\leq a_nf_1(a^\star)+b_nf_2(a^\star)+\varepsilon b_n$ for sufficiently large n and all $a\in(0,1)$.*

Proof. Fix a $\delta > 0$. First consider the situation when $|a - a^\star| \geq \delta$. Let the minimum value of $f_1(a^\star) - f_1(a)$ on the compact set $\{|a - a^\star| \geq \delta\} \cap [0,1]$ be $D > 0$ (D must be positive because f_1 has a unique maximum), and the minimum value of $f_2(a^\star) - f_2(a)$ on $[0,1]$ must be M. Then

$$\begin{aligned} a_n f_1(a^\star) + b_n f_2(a^\star) - a_n f_1(a) - b_n f_2(a) &= a_n\Big(f_1(a^\star) - f_1(a) + \frac{b_n}{a_n}(f_2(a^\star) - f_2(a))\Big) \\ &\geq a_n\Big(D + \frac{b_n}{a_n} M\Big) \geq 0 \end{aligned} \tag{3.33}$$

for large enough n. If $|a - a^\star| < \delta$, then we use $f_1(a) \leq f_1(a^\star)$ and continuity of f_2, and choose δ small enough. □

Since $a \mapsto g(a) - \frac{1}{3}a$ is uniquely maximized at $a^\star$, we can use Lemma 3.9 with $a_n = n \log n$ and $b_n = n$ to obtain the following upper bound, valid for large enough n:

$$\begin{aligned} n \exp\Big(n \log n\,(g(a^\star) - \tfrac{1}{3}a^\star + \varepsilon) - n(1-a^\star)\log(1-a^\star) - (\tfrac{1}{3}a^\star n)\log(\tfrac{1}{3}a^\star n) \\ - \tfrac{2}{3}a^\star n - \tfrac{1}{3}a^\star n \log 6 + a^\star n \log\left(\mathrm{e}^t \lambda\right) + o(n)\Big). \end{aligned} \tag{3.34}$$

Obtaining a lower bound is easy, namely,

$$\mathbb{E}_{p_n(t)}\left[\mathrm{e}^{n \log n\, g(\frac{V_T(G_n)}{n})}\right] \geq \mathrm{e}^{n \log n\, g(a^\star)}\, \mathbb{P}_{p_n(t)}(V_T(G_n) = a^\star n). \tag{3.35}$$

We use the estimates in Theorem 2.1 to obtain the following lower bound for the exponential in (3.34):

$$\begin{aligned} \exp\Big(n \log n\,(g(a^\star) - \tfrac{1}{3}a^\star) - n(1-a^\star)\log(1-a^\star) - (\tfrac{1}{3}a^\star) n \log(\tfrac{1}{3}a^\star) \\ - \tfrac{2}{3}a^\star n - \tfrac{1}{3}a^\star n \log 6 + a^\star n \log \lambda + o(n)\Big). \end{aligned} \tag{3.36}$$

Therefore we obtain the sandwich

$$\begin{aligned} \frac{1}{n}\left(\phi_n(t)\right)^{\binom{n}{2}} \left(\frac{p_n(t)}{p_n}\right)^{a^\star n} \mathrm{e}^{-\varepsilon n} &\leq \frac{1}{Z_n(g)}\, \mathbb{E}_{p_n(t)}\left(\exp\left(n \log n\, g(\tfrac{V_T(G_n)}{n})\right)\right) \\ &\leq n\left(\phi_n(t)\right)^{\binom{n}{2}} \left(\frac{p_n(t)}{p_n}\right)^{a^\star n} \mathrm{e}^{\varepsilon n}, \end{aligned} \tag{3.37}$$

from which the claim follows. □

4. Consistent parameter estimation in exponential random graphs

In practice, we can only observe a large network without knowing its full architecture. From the modeling perspective it is important to be able to estimate unknown parameter(s) from observations. In this section, we show that it is possible to consistently estimate the parameters in the exponential random graph in (3.1), with $\beta = \frac{1}{3} + \frac{\theta}{\log n}$. Note that the distribution of the graph is characterized by two parameters: θ and λ. The estimation procedure is a by product of Theorem 3.1, and is stated in the following theorem:

Theorem 4.1. *Let G_n be an observation from the exponential random graph in (3.1), with $\beta = \frac{1}{3} + \frac{\theta}{\log n}$. Define $\hat{\lambda}_n = \frac{2E(G_n)}{n} - \frac{2V_T(G_n)}{n}$ and*

$$\hat{\theta}_n = \{x\colon\ \arg\max_{0 \leq a \leq 1} \Lambda(a, x, \hat{\lambda}_n) = V_T(G_n)/n\}.$$

Then $\hat{\lambda}_n \xrightarrow{\mathbb{P}_\beta} \lambda$ and $\hat{\theta}_n \xrightarrow{\mathbb{P}_\beta} \theta$ as $n \to \infty$.

Proof. The proof follows from (3.5)–(3.6) and the continuous mapping theorem. □

Remark 4.2. It is not hard to come up with models where consistent estimation is possible via Theorem 3.2. As a proof of concept, let us consider the example in Corollary 3.3. In this case, assume that α is known, and β and λ are unknown parameters with the restriction $3\alpha\beta > 1$. In the notation of Theorem 3.2, $a^\star = \left(\frac{1}{3\alpha\beta}\right)^{\frac{1}{1-\alpha}}$. Therefore a natural proposal for estimators of β and λ is the solution of the equations

$$\left(\frac{1}{3\alpha\hat{\beta}_n}\right)^{\frac{1}{1-\alpha}} = \frac{V_T(G_n)}{n}, \qquad \left(\frac{1}{3\alpha\hat{\beta}_n}\right)^{\frac{1}{1-\alpha}} + \frac{\hat{\lambda}_n}{2} = \frac{2E(G_n)}{n}. \tag{4.1}$$

Using (3.9)–(3.10) and the continuous mapping theorem, we have $\hat{\beta}_n \to \beta$ and $\hat{\lambda}_n \to \lambda$ as $n \to \infty$ under the measure described in Corollary 3.3.

5. Discussion and open problems

In this section, we discuss our main results and list some open problems.

Discussion main results. It is crucial that our main technical theorem, Theorem 2.1, identifies the *second-order asymptotics* of the log of the large deviation probability that $V_T(G_n) \geq an$. The first-order asymptotics, of order $n \log n$, was already identified in [3]. The fact that we can also identify the second-order asymptotics of order n allows us to prove a large deviation principle for the number of edges in the graph (in Theorem 2.3), as well as prove that most triangles are actually vertex disjoint (in Theorem 2.2), which would not have been possible with the first-order result only. This is a reflection of the fact that the key large deviation rate is n, not $n \log n$, as one might have conjectured after [3].

The above results in turn allowed us to suggest a range of *sparse exponential random graph models*, which is important because it is hard to identify sparse exponential random graph models with many triangles. Finally, our results allowed us to prove that the parameters of the model can be consistently estimated, a property that is relevant in practice.

Open problems. Several natural and interesting extensions are possible. The large deviation principle for the number of edges in the graph in Theorem 2.3 suggests that also a *central limit theorem* should also hold. This is further exemplified by the nice limiting generating function for the number of edges in Lemma 2.8. Unfortunately, while being suggestive, it seems hard to turn this observation into a mathematical proof. Indeed, typically some correlation-type inequality is needed for such proofs, which we do not have at our disposal here.

Furthermore, the combination of Theorem 2.2 (describing that most triangles are disjoint) and Theorem 2.3 (describing the convergence of the number of edges) suggests that we might be able to identify the *local limit* of the model as well. Indeed, we conjecture that the local limit is exactly the same as that of the model where $\frac{1}{3}an$ disjoint triangles are randomly dropped inside an Erdős–Rényi random graph. Unfortunately, it seems difficult to prove such a result. Since the degree distribution is uniformly integrable, the model is *tight* in the space of rooted graphs in the local topology, but we do not see how to prove that the limit indicated above really is the local limit.

The type of *models* we investigated may be extended as well. We focussed on the number of vertices in triangles, but it would be natural to consider the number of *edges in triangles* instead. Since this number can vary much more (the number is at most $n(n-1)/2$ rather than n, as for the number of vertices in triangles), it appears to be a significantly more difficult problem. Finally, of course, we could extend the number of parameters in our model, and investigate the behaviour of the

associated exponential random graph. In what generality can the parameters still be consistently estimated?

Acknowledgments. The work in this paper was supported by the Netherlands Organisation for Scientific Research (NWO) through Gravitation-grant NETWORKS-024.002.003.

References

[1] Shankar Bhamidi, Guy Bresler, and Alan Sly. Mixing time of exponential random graphs. *Ann. Appl. Probab.*, 21.
[2] Shankar Bhamidi, Suman Chakraborty, Skyler Cranmer, and Bruce Desmarais. Weighted exponential random graph models: Scope and large network limits. *Journal of Statistical Physics*, 173(3):704–735, 2018.
[3] Suman Chakraborty, Remco van der Hofstad, and Frank den Hollander. Sparse random graphs with many triangles. Preprint arXiv: 2112.06526 [math.PR].
[4] Sourav Chatterjee and Amir Dembo. Nonlinear large deviations. *Advances in Mathematics*, 299:396–450, 2016.
[5] Sourav Chatterjee and Persi Diaconis. Estimating and understanding exponential random graph models. *The Annals of Statistics*, 41(5):2428–2461, 2013.
[6] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume **38** of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.
[7] Ove Frank and David Strauss. Markov graphs. *Journal of the American Statistical association*, 81(395):832–842, 1986.
[8] Olle Häggström and Johan Jonasson. Phase transition in the random triangle model. *Journal of Applied Probability*, 36(4):1101–1115, 1999.
[9] Mark S Handcock. Assessing degeneracy in statistical models of social. 2003.
[10] Paul W Holland and Samuel Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association*, 76(373):33–50, 1981.
[11] David R Hunter and Mark S Handcock. Inference in curved exponential family models for networks. *Journal of computational and graphical statistics*, 15(3):565–583, 2006.
[12] Johan Jonasson. The random triangle model. *Journal of Applied Probability*, 36(3):852–867, 1999.
[13] Sumit Mukherjee. Degeneracy in sparse ERGMs with functions of degrees as sufficient statistics. *Bernoulli*, 26(2):1016–1043, 2020.
[14] Mark EJ Newman. Random graphs with clustering. *Physical Review Letters*, 103(5):058701, 2009.
[15] Anatol Rapoport. Cycle distributions in random nets. *The Bulletin of Mathematical Biophysics*, 10(3):145–157, 1948.
[16] Nathan Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011.
[17] M Ángeles Serrano and Marian Boguña. Clustering in complex networks. I. General formalism. *Physical Review E*, 74(5):056114, 2006.
[18] Tom AB Snijders, Philippa E Pattison, Garry L Robins, and Mark S Handcock. New specifications for exponential random graph models. *Sociological methodology*, 36(1):99–153, 2006.
[19] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

Email address: contact@sumanc.com

Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, Netherlands
Email address: rhofstad@win.tue.nl

Mathematical Institute, Leiden University, Leiden, Netherlands
Email address: denholla@math.leidenuniv.nl