The Annals of Applied Probability
2011, Vol. 21, No. 6, 2146–2170
DOI: 10.1214/10-AAP740

MIXING TIME OF EXPONENTIAL RANDOM GRAPHS

BY SHANKAR BHAMIDI[1], GUY BRESLER[2] AND ALLAN SLY[3]

University of North Carolina, Chapel, University of California, Berkeley and Microsoft Research, Redmond

A variety of random graph models has been developed in recent years to study a range of problems on networks, driven by the wide availability of data from many social, telecommunication, biochemical and other networks. A key model, extensively used in sociology literature, is the exponential random graph model. This model seeks to incorporate in random graphs the notion of reciprocity, that is, the larger than expected number of triangles and other small subgraphs. Sampling from these distributions is crucial for parameter estimation hypothesis testing and more generally for understanding basic features of the network model itself. In practice, sampling is typically carried out using Markov chain Monte Carlo, in particular, either the Glauber dynamics or the Metropolis–Hastings procedure.

In this paper we characterize the high and low temperature regimes of the exponential random graph model. We establish that in the high temperature regime the mixing time of the Glauber dynamics is $\Theta(n^2 \log n)$, where n is the number of vertices in the graph; in contrast, we show that in the low temperature regime the mixing is exponentially slow for any local Markov chain. Our results, moreover, give a rigorous basis for criticisms made of such models. In the high temperature regime, where sampling with Markov chain Monte Carlo is possible, we show that any finite collection of edges is asymptotically independent; thus, the model does not possess the desired reciprocity property and is not appreciably different from the Erdős–Rényi random graph.

1. Introduction. In the recent past there has been an explosion in the study of real-world networks including rail and road networks, biochemical networks, data communication networks such as the internet and social networks. This has resulted in a concerted interdisciplinary effort to develop new mathematical network models to explain characteristics of observed real world networks, such as power law degree behavior, small world properties and a high degree of clustering (see e.g., [1, 7, 15] and the citations therein).

Received December 2008; revised October 2010.

[1]Supported by NSF Grant DMS-07-04159, by PIMS and NSERC Canada and by a UNC Chapel Hill Research council grant.

[2]Supported by a Vodafone Fellowship and an NSF Graduate Fellowship.

[3]Supported by NSF Grants DMS-05-28488 and DMS-05-48249 and by DOD ONR Grant N0014-07-1-05-06.

MSC2010 subject classifications. 60C05, 05C80, 90B15.

Key words and phrases. Mixing times, exponential random graphs, path coupling.

Clustering (or reciprocity) refers to the prevalence of triangles in a graph. This phenomenon is most easily motivated in social networks, where nodes represent people and edges represent relationship. The basic idea is that if two individuals share a common friend, then they are more likely than otherwise to themselves be friends. However, most of the popular modern network models, such as the preferential attachment and the configuration models, are essentially tree-like and thus do not model the reciprocity observed in real social networks.

One network model that attempts to incorporate reciprocity is the exponential random graph model. This model is especially popular in the sociology community. The model follows the statistical mechanics approach of defining a Hamiltonian to weight the probability measure on the space of graphs, assigning higher mass to graphs with "desirable" properties. One justification for using this approach to defining a probability distribution is the so-called Maximum Entropy Principle, introduced by Jaynes [11, 12]. The principle says to select from among a set of candidate distributions the one with highest entropy, which then results in an exponential distribution of the form below when the set of distributions in question satisfies certain constraints. While deferring the general definition of the model to Section 1.1, let us give a brief example. Fix parametric constants $h, \beta > 0$ and for every graph X on n labeled vertices with $E(X)$ edges and $T(X)$ triangles, define the Hamiltonian of the graph as

$$H(X) = hE(X) + \beta T(X).$$

A probability measure on the space of graphs may then be defined as

$$p_n(X) = \frac{e^{H(X)}}{Z}, \tag{1}$$

where Z is the normalizing constant, often called the partition function. More generally, one can consider Hamiltonians in graphs which include counts $T_i(X)$ of different small subgraphs G_i,

$$H(X) = \sum_i \beta_i T_i(X).$$

For a nice discussion of the implications of the formulation of this model in statistics, see [13].

Social scientists use these models in several ways. The class of distributions (1) is an exponential family which allows for statistical inference of the parameters using the subgraph counts (which are sufficient statistics for the parameters involved). Sociologists carry out tests of significance hoping to understand how prescription of local quantities, such as the typical number of small subgraphs in the network, affects more global macroscopic properties. Parameter estimation can be carried out either by maximum likelihood or, as is more commonly done, by simply equating the subgraph counts. Both procedures generally require sampling, in the case of maximum likelihood, to estimate the normalizing constants. Thus,

efficient sampling techniques are key to statistical inference on such models. At a more fundamental level, sociologists are interested in the the question of how localized phenomena involving a small number people determine the large scale structure of the networks [18]. Sampling exponential random graphs and observing their large scale properties is one way this can be realized. Sampling is almost always carried out using local Markov chain Monte Carlo (MCMC) algorithms, in particular, the Glauber dynamics or Metropolis–Hastings. These are reversible ergodic Markov chains which eventually converge to the stationary distribution $p_n(X)$. However, our results show that the time to convergence can vary enormously depending on the choice of parameters.

Our results. It is surprising that, in spite of the practical importance of sampling from exponential random graph distributions, there has been no mathematically rigorous study of the mixing time of any of the various Markov chain algorithms in this context. The goal of this paper is to fill this gap. We focus attention on the Glauber dynamics, one of the most popular Markov chains. We provide the first rigorous analysis of the mixing time of the Glauber dynamics for the above stationary distribution and do so in a very general setup. In the process we give a rigorous definition of the "high temperature" phase, where the Gibbs distribution is unimodal and the Glauber dynamics converges quickly to the stationary distribution and the "low temperature" phase, where the Gibbs distribution is multimodal and the Glauber dynamics takes an exponentially long time to converge to the stationary distribution. While a complete understanding of the Gibbs distribution in the low temperature phase remains out of reach (see, however, the important work of Chatterjee and Varadhan in the case of triangles [5]), we can nevertheless show that the distribution has poor conductance, thereby establishing exponentially slow mixing for any local Markov chain with the specified stationary distribution.

Our results, moreover, give a rigorous basis for criticisms made of such models. In the high temperature regime, where sampling with MCMC is possible, we show that any finite collection of edges are asymptotically independent. Also, we show that with exponentially high probability, a sampled graph is *weakly pseudo-random*, meaning that it satisfies a number of equivalent properties (such as high edge expansion) shared by Erdős–Rényi random graphs. Thus, the model does not possess the desired reciprocity property and is not appreciably different from the Erdős–Rényi random graph.

Related literature. There is a large body of literature, especially in the social networking community, on exponential random graph models. We shall briefly mention just some of the relevant literature and how it relates to our results (see [3, 15, 18] and the references therein for more background). The pioneering article in this area by Frank and Strauss [10] introduced the concept of Markov graphs. Markov graphs are a special case of exponential random graphs where the subgraphs are stars or triangles. Extending the methodology of [10], Wasserman and

Pattison [19] introduced general subgraph counts. However, from the outset a number of researchers noted problems at the empirical level for their Markov chain algorithms, depending on parameter values. See [18] for a relevant discussion of empirical findings as well as several new specifications of the model to circumvent such issues.

On the theoretical side, Chatterjee and Varadhan [5], in their recent work characterizing the large deviation properties of Erdős–Rényi random graphs, developed mathematical techniques that can be used to study the distribution of these random graphs. At the statistical physics (nonrigorous) level Newman and his co-authors have studied the case where the subgraphs are triangles and 2-stars. In this setting, using mean-field approximations, they predicted a phase transition between a high-symmetry phase, with graphs exhibiting only a mild amount of reciprocity, and a degenerate symmetry-broken phase with either high or low edge density (see [16] and [17]).

1.1. *Definitions and notation.* This section contains a precise mathematical definition of the model and the Markov chain methodology used in this paper. We work on the space $\mathcal{G}_n$ of all graphs on n vertices with vertex set $[n] := \{1, 2, \ldots, n\}$. The edges are assumed to be undirected. We shall use $X = (x_e)$ to denote a graph from $\mathcal{G}_n$ where for every (undirected) edge $e = (i, j)$, x_e is 1 if the edge between vertex i and j is present and 0 otherwise. For simplicity, we shall often write $X(e)$ for x_e. Here and occasionally later in the paper, the term "edge" is used to mean a pair of vertices, with x_e encoding whether the edge is actually present in the graph. The exponential random graph model is defined in terms of the number of subgraphs G (e.g., triangles or edges) contained in X. It will be convenient to define these subgraph counts as follows. Fix a graph G on the vertex set $1, 2, \ldots, m$. Let $([n])_m$ denote the set of all m tuples of distinct elements,

$$([n])_m := \{(v_1, \ldots, v_m) : v_i \in [n], v_1 \neq v_2 \neq \cdots \neq v_m\}.$$

We shall denote such an m tuple of distinct vertices by $\mathbf{v}_m$. In a graph X, for any m distinct vertices $\mathbf{v}_m$, let $H_X(\mathbf{v}_m)$ denote the subgraph of X induced by $\mathbf{v}_m$. Say that $H_X(\mathbf{v}_m)$ contains G, denoted by $H_X(\mathbf{v}_m) \cong G$, if whenever the edge (i, j) is present in G, then the edge (v_i, v_j) is present in $H_X(\mathbf{v}_m)$ for all $\{1 \leq i \neq j \leq m\}$. For a configuration $X \in \mathcal{G}_n$ and a fixed graph G, define the count

$$N_G(X) = \sum_{\mathbf{v}_m \in ([n])_m} \mathbb{1}\{H_X(\mathbf{v}_m) \cong G\}. \tag{2}$$

This definition is equivalent to the usual exponential random graph model up to adjustments in the constants β by multiplicative factors. It counts subgraphs multiple times; for instance, a triangle will be counted 6 times and in general a graph G with k automorphisms will be counted k times. By dividing the parameters β_i by this multiplicative factor we reduce to the usual definition.

In our proof we shall also need more advanced versions of the above counts which we define now. Fix an edge $e = (a, b) \in X$. The subgraph count of G in $X \cup \{e\}$ containing edge e is defined as

$$N_G(X, e) = \sum_{\mathbf{v}_m \in ([n])_m, \mathbf{v}_m \ni a,b} \mathbb{1}\{H_{X\cup\{e\}}(\mathbf{v}_m) \cong G\}.$$

Here, notation is abused slightly, with $\mathbf{v}_m \ni a, b$ indicating that the tuple $\mathbf{v}_m = (v_1, \ldots, v_m)$ has $v_i = a$ and $v_j = b$ for some $1 \le i, j \le m$. Similarly, for two edges $e = (a, b)$ and $e' = (c, d)$, define the subgraph counts of G in $X \cup \{e, e'\}$ and containing edges e, e' by

$$N_G(X, e, e') = \sum_{\mathbf{v}_m \in ([n])_m, \mathbf{v}_m \ni a,b,c,d} \mathbb{1}\{H_{X\cup\{e,e'\}}(\mathbf{v}_m) \cong G\}.$$

Gibbs measure. We now define the probability measure on the space $\mathcal{G}_n$. Fix $s \ge 1$ and fix graphs $G_1, G_2, \ldots, G_s$ with G_i a graph on $|V_i|$ labeled vertices, with $|V_i| \le L$ and with edge set E_i. For simplicity we shall think of G_i as a graph on the vertex set $1, 2, \ldots, |V_i|$. By convention we shall always let G_1 denote the edge graph consisting of the graph with vertex set $1, 2$ and edge set $(1, 2)$. In this notation, for any configuration $X \in \mathcal{G}_n$, the quantity $N_{G_1}(X)$ will be twice the number of edges in X. With this convention, fix constants $\beta_1, \beta_2, \ldots, \beta_s$ with $\beta_i > 0$ for $i \ge 2$ and $\beta_1 \in \mathbb{R}$. The exponential random graph probability measure is defined as follows.

DEFINITION 1. For $G_1, \ldots, G_s$ and constants $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_s)$ as above, the Gibbs measure on the space $\mathcal{G}_n$ is defined as the probability measure

$$p_n(X) = \frac{1}{Z_n(\boldsymbol{\beta})} \exp\left(\sum_1^s \beta_i \frac{N_{G_i}(X)}{n^{|V_i|-2}}\right), \qquad X \in \mathcal{G}_n. \tag{3}$$

Here $Z_n(\boldsymbol{\beta})$ is the normalizing factor and is often called the partition function. For simplicity we have suppressed the dependence of the measure on the vector $\boldsymbol{\beta}$. We note that since β is constant, the number of edges is of the order n^2 and in general, the subgraph count $N_{G_i}(X)$ is of the order $n^{|V_i|}$. Thus, the subgraph count N_{G_i} is normalized by the factor $n^{|V_i|-2}$ so that the contribution of each factor scales properly and is of order n^2 in the large n limit. Setting $\beta_i \ge 0$ for $i \ge 2$ makes the Gibbs measure a monotone (also ferromagnetic) system which will be important for our proof. The term β_1 does not affect the interaction between edges and plays the role of an external field in this model; adjusting β_1 makes it more or less likely for edges to be included.

The term in the exponent is often called the *Hamiltonian* and we shall denote it by

$$H(X) = \sum_1^s \beta_i \frac{N_{G_i}(X)}{n^{|V_i|-2}}.$$

Note that $H(X):\{0,1\}^{\binom{n}{2}} \to \mathbb{R}^+$ is a function of $\binom{n}{2}$ Boolean variables $X(e)$ and has an elementary Fourier decomposition in terms of the basis functions $\prod_{e\in S} X(e)$, where S runs over all possible subsets of edges. Thus, with respect to any fixed edge e, we can decompose the above Hamiltonian as

$$H(X) = A_e(X) + B_e(X),$$

where A_e consists of all terms dependent on edge e and $B_e(X)$ denotes all terms independent of edge e. Let X_{e+} denote the configuration of edges which coincides with X for edges $e \neq f$ and has $X_{e+}(e) = 1$. The partial derivative with respect to the edge e of the Hamiltonian H, evaluated at a configuration X, is defined by the formula

$$\partial_e H(X) = A_e(X_{e+}).$$

The higher derivatives $\partial_e \partial_{e'}$ for $e \neq e'$ are defined similarly by iterating the above definition.

Partial ordering on configurations. We note that the configuration space $\mathcal{G}_n$ of simple graphs on n vertices can be partially ordered in the sense that for two configurations $X, Y \in \mathcal{G}_n$, say that $X \leq Y$ is every edge in X is also in Y.

In passing, we note that in the above model specification, the parameters β_i are kept fixed and nonvarying in n so that the Gibbs measure in equation (3) results in graphs which are *dense* in the sense that they typically tend to have $\Theta_P(n^2)$ edges. One could formulate models where the $\beta_i = \beta_i(n)$ are functions of n and, for example, taking $\beta_1 \to -\infty$ as $n \to \infty$ at an appropriate rate, end up with *sparse* random graph models [models which have $\Theta_P(n)$ edges]. We shall not pursue this line of inquiry in this article and defer this to a later study.

Glauber dynamics and local chains. The Glauber dynamics is an ergodic reversible Markov chain with stationary distribution $p_n(\cdot)$, where at each stage exactly one edge is updated. It is defined as follows:

DEFINITION 2. Given the Gibbs measure stated above, the corresponding Glauber dynamics is a discrete time ergodic Markov chain on $\mathcal{G}_n$. Given the current state X, the next state X' is obtained by choosing an edge (i.e., pair of vertices) e uniformly at random and letting $X' = X_{e+}$ with probability proportional $p_n(X_{e+})$ and $X'(e) = X_{e-}$ with probability proportional to $p_n(X_{e-})$. Here X_{e+} is the graph which coincides with X for all edges other than e and $X_{e+}(e) = 1$. Similarly, X_{e-} is the graph which coincides with X for all edges other than e and $X_{e-}(e) = 0$.

There are various other chains that can also be used to simulate the above Gibbs measure. Call a chain on $\mathcal{G}_n$ *local* if at most $o(n)$ edges are updated in each step. The transition rates for the Glauber dynamics satisfy the following relation:

LEMMA 3. *Given that we chose edge e to update, the probability of the transition $X \hookrightarrow X_{e+}$ is $\frac{\exp(\partial_e H(X))}{1+\exp(\partial_e H(X))}$ and the probability of the transition $X \hookrightarrow X_{e-}$ is $\frac{1}{1+\exp(\partial_e H(X))}$.*

Mixing time. We will be interested in the time it takes for the Glauber dynamics to get close to the stationary distribution given by the Gibbs measure (3). The *mixing time* τ_{mix} of a Markov chain is defined as the number of steps needed in order to guarantee that the chain, starting from an arbitrary state, is within total variation distance e^{-1} from the stationary distribution.

We mention the following fundamental result which draws a connection between total variation distance and coupling. It allows us to conclude that if we can couple two versions of the Markov chains started from different states quickly, the chain mixes quickly. The following lemma is well known (see e.g., [2]).

LEMMA 4 (Mixing time lemma). *For a Markov chain X, suppose there exist two coupled copies, Y and Z, such that each is marginally distributed as X and*

$$\max_{y,z} \mathbf{P}(Y_{t_0} \neq Z_{t_0} | Y_0 = y, Z_0 = z) \leq e^{-1}.$$

Then the the mixing time of X satisfies $\tau_{\text{mix}} \leq t_0$.

Recall from Section 1.1 that the configuration space of graphs $\mathcal{G}_n$ can be partially ordered. Since the exponential random graph model is a monotone system, we can couple the Glauber dynamics so that if $X(0) \leq Y(0)$, then for all t, $X(t) \leq Y(t)$. This inequality is a partial ordering meaning that the edge set of X is a subset of the edge set of Y. This is known as the monotone coupling and, by monotonicity, Lemma 4 reduces to bounding the time until chains starting from the empty and complete graphs couple.

With the above definitions of the Gibbs measure, the following functions determine the properties of the mixing time. Define for fixed $\boldsymbol{\beta} \in \mathbb{R} \times (\mathbb{R}_+)^{s-1}$ the functions

$$\Psi_{\boldsymbol{\beta}}(p) = \sum_{i=1}^{s} 2\beta_i |E_i| p^{|E_i|-1},$$

where we recall that E_i are the fixed graphs in the definition of the Gibbs measure. Now define the function

$$\varphi_{\boldsymbol{\beta}}(p) = \frac{\exp(\Psi_{\boldsymbol{\beta}}(p))}{1 + \exp(\Psi_{\boldsymbol{\beta}}(p))}.$$

Note that $\Psi_{\boldsymbol{\beta}}$ is a smooth, strictly increasing function on the unit interval. Since $\varphi_\beta(0) > 0$ and $\varphi_\beta(1) < 1$, the equation $\varphi_\beta(p) = p$ has at least one solution, denoted by p^*. If this solution is unique and not an inflection point, then $0 < \varphi'_\beta(p^*) < 1$. The function $\varphi(p)$ has the following loose motivation: if X is a graph chosen according to the Erdős–Rényi distribution $G(n, p)$, then, with high probability, all edge update probabilities $\frac{\exp(\partial_e H(X))}{1+\exp(\partial_e H(X))}$ are approximately $\varphi(p)$.

Phase identification. We now describe the high and low temperature phases of this model. Recall that our parameter space is $\mathcal{B} = \mathbb{R} \times (\mathbb{R}_+)^{s-1}$. We call $p \in [0, 1]$ a fixed point if $\varphi_{\boldsymbol{\beta}}(p) = p$.

High temperature phase. We say that a $\beta \in \mathcal{B}$ belongs to the high temperature phase if $\varphi_\beta(p) = p$ has a unique fixed point p^* which satisfies

$$\varphi'_\beta(p^*) < 1. \tag{4}$$

Low temperature phase. We say that a $\beta \in \mathcal{B}$ belongs to the low temperature phase if $\varphi_\beta(p) = p$ has at least two fixed points p^* which satisfy $\varphi'_\beta(p^*) < 1$.

Values of β not in either phase are said to be in the critical points. They occur when one of the fixed points is an inflection point of φ_β. These critical points form an $s-1$ dimensional manifold which is in the intersection of the closure of the high and low temperature phases. We do not consider the critical points in this paper. For simplicity, in the proof we shall suppress the dependence of the functions on $\boldsymbol{\beta}$ and write φ for $\varphi_{\boldsymbol{\beta}}$ and Ψ for $\Psi_{\boldsymbol{\beta}}$. The dependence on β is implicit also in the $O(\cdot)$ and $\Theta(\cdot)$ notation for the mixing times.

1.2. *Results.* The first two results show that the high and low temperature phases determine the mixing time for local Markov chains.

THEOREM 5 (High temperature). *If $\varphi(p)$ is in the high temperature regime then the mixing time of the Glauber dynamics is $\Theta(n^2 \log n)$.*

THEOREM 6 (Low temperature). *If $\varphi(p)$ is in the low temperature regime then the mixing time of the Glauber dynamics is $e^{\Omega(n)}$. Furthermore, this holds not only for the Glauber dynamics but for any local dynamics on $\mathcal{G}_n$.*

The next theorem shows that the exponential random graph model is not appreciably different from Erdős–Rényi random graph model in the high temperature regime where sampling is possible.

THEOREM 7 (Asymptotic independence of edges). *Let X be drawn from the exponential random graph distribution in the high temperature phase. Let $e_1, \dots, e_k$ be an arbitrary collection of edges with associated indicator random variables $x_{e_i} = \mathbb{1}(e_i \in X)$. Then for all $(a_1, \dots, a_k) \in \{0, 1\}^k$, the random variables $x_{e_1}, \dots, x_{e_k}$ satisfy*

$$\left|\mathbf{P}(x_1 = a_1, \dots, x_k = a_k) - (p^*)^{\sum a_i}(1 - p^*)^{k - \sum a_i}\right| \to 0$$

as $n \to \infty$. Thus, the random variables $x_{e_1}, \dots, x_{e_k}$ are asymptotically independent.

A consequence is that a graph sampled from the exponential random graph distribution is with high probability weakly pseudo-random (see [14] or [6]). This means that it satisfies a number of equivalent properties, including large spectral gap and correct number of subgraph counts, that make it very similar to an Erdős–Rényi random graph.

COROLLARY 8 (Weak pseudo-randomness). *With probability* $1 - o(1)$ *an exponential random graph is weakly pseudo-random.*

1.3. *Idea of the proof.* We give a summary of the main ideas of the proof:

1. Consider first the high temperature phase. A natural approach to bounding the coupling time and hence, the mixing time by Lemma 4, is to use the technique of *path coupling* [4]. In path coupling, instead of trying to couple from every pair of states, we try to show that for any pair of states x and y that differ in a single edge there exists a coupling of two copies of the chain started at x and y such that

$$\mathbf{E}\big(d_H(X(1), Y(1)) | X(0) = x, Y(0) = y\big) \le (1 - \gamma) \tag{5}$$

   for some $\gamma = \gamma(n)$, where d_H is the Hamming distance. However, this approach fails for some φ_β in the high temperature regime when $\sup_{0 \le p \le 1} \varphi'(p) > 1$.
2. It turns out that the configurations in the high temperature regime where path coupling fails are very rare under the Gibbs measure. We therefore define a set [a neighborhood of the unique fixed point $\varphi'_{\boldsymbol{\beta}}(\cdot) < 1$] in which path coupling does give a contraction. More precisely, for a configuration X, define

$$r_G(X, e) = \left(\frac{N_G(X, e)}{2|E|n^{|V|-2}} \right)^{1/(|E|-1)}. \tag{6}$$

   This is (asymptotically) the maximum likelihood choice for the parameter p of the Erdős–Rényi random graph on n vertices, $G(n, p)$, having observed $N_G(X, e)$ subgraphs G containing the edge e. Let $\mathbb{G}_L$ denote the class of all graphs with at most L vertices, where L is some integer greater than or equal to $\max_i |V_i|$, that is, the maximum number of vertices in the graphs fixed in the definition of the Gibbs measure. What we prove is that for ε small enough, if the two configurations x and y belong to the set

$$\mathbf{G} := \Big\{ X : \max_{\substack{G \in \mathbb{G}_L \\ e \in E}} |r_G(X, e) - p^*| < \varepsilon \Big\},$$

   then equation (5) holds for $\gamma(n) = \delta/n^2$ for some $\delta > 0$. Thus, starting from any state x, if we can show that in a small number of steps [$O(n^2)$ is enough] we reach $\mathbf{G}$, then a variant of path coupling proves rapid mixing. This preliminary stage where we run the Markov chain for some steps so that it reaches a "good configuration" is termed the *burn in* phase. This approach has been used before, particularly in proving mixing times for random colorings, for example, in [8].

3. To show that we enter the good set $\mathbf{G}$ quickly, we control all the $r_G(X, e)$ for all subgraphs $G \in \mathbb{G}_L$ simultaneously and via a coupling with biased random walks show that with exponentially high probability for large n, within $O(n^2)$ steps we reach the set $\mathbf{G}$. We crucially make use of the monotonicity of the system here by writing the drifts in terms of the $r_G(X, e)$ and bounding them by their maximum. This completes the proof for the rapid mixing in the high temperature phase. This also shows how in the high temperature phase, most of the Gibbs measure of the exponential random graph model is concentrated on configurations which are essentially indistinguishable from the Erdős–Rényi $G(n, p^*)$ random graph model.
4. In the low temperature phase we use a conductance argument to show slow mixing for any Markov chain that updates $o(n)$ edges per time step. The argument makes use of the same random walk argument used in the burn in stage to bound the measure of certain sets of configurations under the Gibbs measure. Specifically, we show that for every fixed point p^* of the equation $\varphi(p) = p$ with $\varphi'(p) < 1$, the Glauber dynamics allows an exponentially small flow of probability to leave the set of configurations that are nearly indistinguishable from an Erdős–Rényi random graph with parameter p^*. Because the stationary distribution of the Glauber dynamics is the Gibbs measure, this allows us to bound the relative measure of the sets under consideration thereby showing that if we have two or more fixed points p^*, then it takes an exponentially long time for configurations to leave the set of configurations indistinguishable from an Erdős–Rényi random graph with parameter p^*. Thus, mixing takes an exponentially long time.

2. Proof of the main results.

2.1. *Subgraph counts.* Before starting the proof we need a couple of simple lemmas on the subgraph counts. For a graph $X \in \mathcal{G}_n$ recall the subgraph counts $N_G(X)$ of a predefined graph G on m nodes as well as the counts in X of the subgraphs containing edges, namely, $N_G(X, e)$ and $N_G(X, e, e')$ as defined in Section 1.1.

The following lemma records the quantities $N_G(X)$, $N_G(X, e)$ and $N_G(X, e, e')$ for the complete graph $X = K_n$.

LEMMA 9. *Consider the complete graph on n vertices K_n and let $N_G(K_n)$, $N_G(K_n, e)$ and $N_G(K_n, e, e')$ be defined as above. Then*:

(a)
$$N_G(K_n) = \binom{n}{|V|} |V|! \sim n^{|V|}.$$

(b)
$$N_G(K_n, e) = 2|E| \binom{n-2}{|V|-2} (|V|-2)! \sim 2|E| \cdot n^{|V|-2}.$$

(c) *For a fixed edge e we have*

$$\sum_{e' \neq e} N_G(K_n, e, e') = (|E| - 1) N_G(K_n, e) \sim 2|E|(|E| - 1) n^{|V|-2}.$$

LEMMA 10. *For an edge α in the graph G, denote by G_α the graph obtained from G by removing the edge α. Then*

$$\sum_{e' \neq e} N_G(X, e, e') = \sum_{\substack{\alpha \in E(G) \\ \alpha \neq e}} N_{G_\alpha}(X, e). \tag{7}$$

PROOF. The sum on the left-hand side of 7 counts the total number of isomorphic embeddings of G that contain the edge e in the configuration $X \cup \{e'\}$, for some e' with the edge e' marked. Now, each isomorphism with marked edge e' is counted on the right-hand side of 7 for the choice α equal to the marked edge in the graph G, with the same isomorphism restricted to G_α. Conversely, for each $\alpha \in E(G)$ and each subgraph embedding, the same embedding is counted on the left-hand side with the edge e' situated at the location α. □

2.2. *Burn-in period.* In this section we show that after a suitably short "burn-in" period, the Markov chain is in the good set $\mathbf{G}$. We first define the following important construct. Recall that $\mathbb{G}_L$ was the set of all graphs with less than L vertices.

DEFINITION 11. For any graph $X \in \mathcal{G}_n$, define

$$r_{\max}(X) = \max_{e, G_\lambda \in \mathbb{G}_L} r_{G_\lambda}(X, e)$$

and similarly define $r_{\min}(X) = \min_{e, G_\lambda \in \mathbb{G}_L} r_{G_\lambda}(X, e)$.

Lemma 12 bounds the expected drift of $r_{\max}(X)$.

LEMMA 12. *The expected change in $N_G(X, e)$ after one step of the Glauber dynamics, starting from the configuration X, can be bounded as*

$$\mathbf{E}\left[\frac{N_G(X(1), e) - N_G(X(0), e)}{n^{|V|-2}}\right] \leq (1 + o(1)) \frac{2}{\binom{n}{2}} |E|(|E| - 1)\left[-r_G(X, e)^{|E|-1} + \varphi(r_{\max})(r_{\max})^{|E|-2}\right],$$

where for ease of notation we have used $r_{\max} = r_{\max}(X)$ and have suppressed the dependence of this object on the configuration X.

PROOF. The expected change, after one step of the Glauber dynamics, in the number of isomorphisms from G to subgraphs of X containing the edge e can be counted by first negating the expected loss in number when removing a random edge e' (leaving the configuration unchanged if e' was not present) and then adding the expected number of graphs created by including a random edge e'. This gives

$$\begin{aligned}
&\mathbf{E}\left[\frac{N_G(X(1),e)-N_G(X(0),e)}{n^{|V|-2}}\right]\\
&\quad=\frac{1}{n^{|V|-2}}\left[-\binom{n}{2}^{-1}(|E|-1)N_G(X,e)\right.\\
&\qquad\left.+\binom{n}{2}^{-1}\sum_{e'\neq e}N_G(X,e,e')\mathbf{P}\big(X_{e'}(1)=1|e'\text{ updated}\big)\right].
\end{aligned}\tag{8}$$

Now, we may upper bound the probability of including an edge using Lemma 3 and the definition of $r_{\max}$:

$$\begin{aligned}
&\mathbf{P}\big(X_{e'}(1)=1|e'\text{ updated}\big)\\
&\quad=\frac{\exp(\partial_{e'}H(X))}{\exp(\partial_{e'}H(X))+1}\\
&\quad=\exp\Big(\sum_i\beta_i\frac{N_{G_i}(X,e')}{n^{|V|-2}}\Big)\Big/\Big(\exp\Big(\sum_i\beta_i\frac{N_{G_i}(X,e')}{n^{|V|-2}}\Big)+1\Big)\\
&\quad\le\exp\Big(\sum_i\beta_i\frac{N_G(K_n,e')(r_{\max})^{|E_i|-1}}{n^{|V_i|-2}}\Big)\\
&\qquad\Big/\Big(\exp\Big(\sum_i\beta_i\frac{N_G(K_n,e')(r_{\max})^{|E_i|-1}}{n^{|V_i|-2}}\Big)+1\Big)\\
&\quad=\varphi(r_{\max}).
\end{aligned}\tag{9}$$

Next, by Lemmas 9 and 10 and the definition of $r_{\max}$, we have

$$\begin{aligned}
\sum_{e'}N_G(X,e,e')&=\sum_\alpha N_{G_\alpha}(X,e)\\
&\le\sum_\alpha N_{G_\alpha}(K_n,e)(r_{\max})^{|E|-2}\\
&=\sum_{e'\neq e}N_G(K_n,e,e')(r_{\max})^{|E|-2}\\
&=|E|(|E|-1)2n^{|V|-2}(r_{\max})^{|E|-2}\big(1+o(1)\big).
\end{aligned}\tag{10}$$

Using the estimates (9) and (10), equation (8) gives

$$\begin{aligned}
&\mathbf{E}\Big[\frac{N_G(X(1),e)-N_G(X(0),e)}{n^{|V|-2}}\Big]\\
&\quad\le (1+o(1))\frac{1}{n^{|V|-2}\binom{n}{2}}\big[-(|E|-1)N_G(X,e)\\
&\qquad\qquad +\varphi(r_{\max})2|E|(|E|-1)n^{|V|-2}(r_{\max})^{|E|-2}\big]\\
&\quad= (1+o(1))\frac{1}{n^{|V|-2}\binom{n}{2}}\big[-(|E|-1)2|E|n^{|V|-2}r_G(X,e)^{|E|-1}\\
&\qquad\qquad +\varphi(r_{\max})2|E|(|E|-1)n^{|V|-2}(r_{\max})^{|E|-2}\big]\\
&\quad= (1+o(1))\frac{2}{\binom{n}{2}}|E|(|E|-1)\big[-r_G(X,e)^{|E|-1}+\varphi(r_{\max})(r_{\max})^{|E|-2}\big].
\end{aligned}$$

We have the following analogous result for $r_{\min}$. The proof is identical so we do not give the proof. □

LEMMA 13. *The expected change in $N_G(X,e)$ after one step of the Glauber dynamics, starting from the configuration X, can be lower bounded as*

$$\begin{aligned}
&\mathbf{E}\Big[\frac{N_G(X(1),e)-N_G(X(0),e)}{n^{|V|-2}}\Big]\\
&\quad\ge (1+o(1))\frac{2}{\binom{n}{2}}|E|(|E|-1)\big[-r_G(X,e)^{|E|-1}+\varphi(r_{\min})(r_{\min})^{|E|-2}\big],
\end{aligned}$$

where for ease of notation we have used $r_{\min}=r_{\min}(X)$ and have suppressed the dependence of this object on the configuration X.

The following lemma shows that if we start from a configuration where $r_{\max}(X)$ is significantly different from p^* (a solution of the fixed point equation) but bounded away from any other solution of this fixed point equation, then there is a drift of the Glauber dynamics toward a configuration where $r_{\max}(X)$ is closer to p^* than the starting state.

LEMMA 14. *Let p^* be a solution of the equation $\varphi(p)=p$ with $\varphi'(p^*)<1$ and let $\bar{p}$ be the least solution greater than p^* of the equation $\varphi(p)=p$ if such a solution exists or* 1, *otherwise. Let the initial configuration be $X(0)$ with $p^*+\mu\le r_{\max}(X(0))\le\bar{p}-\mu$ for some $\mu>0$. Then there is a $\delta,c>0$ depending only on μ, L and φ, so that after $T=cn^2$ steps of the Glauber dynamics, it holds that $r_{\max}(X(T))\le r_{\max}(X(0))-\delta$ with probability $1-e^{-\Omega(n)}$.*

PROOF. The lemma is proved using a moment generating function argument; it is shown that in the relevant regime, each of the random variables $N_G(X(t), e)$ (one for each edge e and graph G) behaves like a biased random walk.

Choose $\varepsilon, \delta > 0$ so that for any $r \in [p^* + \mu, \bar{p} - \mu - \delta]$,

$$(r - 2\delta)^{|E|-1} > \varphi(r + \delta)(r + \delta)^{|E|-2} + \varepsilon. \tag{11}$$

It follows by Lemma 12 that if $r_G(X(t), e) \geq r_{\max}(X(0)) - 2\delta$ and $r_{\max}(X(t)) \leq r_{\max}(X(0)) + \delta$, then for sufficiently large n,

$$\mathbf{E}\left[\frac{N_G(X(t+1), e) - N_G(X(t), e)}{n^{|V|-2}}\right] \leq -\gamma/n^2$$

for some $\gamma > 0$ depending only on φ, δ and ε. Using this negative drift, we bound the probability that any of the random variables $r_G(X(t), e)$ exceed $r_{\max}(X(0)) + \delta$ before time T.

Define the event

$$A_t(\delta) = \bigcap_{e,G} \{r_G(X(t), e) \leq r_{\max}(X(0)) + \delta\}$$

and put

$$D_t(e, G, \delta) = A_t \cap \{r_{\max}(X(0)) - 2\delta \leq r_G(X(t), e) \leq r_{\max}(X(0)) + \delta\}$$

and

$$B_{t_1,t_2}(e, G, \delta) = \left(\bigcap_{t_1 \leq t < t_2} D_t\right) \cap \{r_G(X(t_2), e) - r_G(X(t_1), e) > \delta/2\}.$$

$B_{t_1,t_2}(e, G, \delta)$ is the event that all the edge statistics $r_{G'}(X(t), e')$ behave well starting at time t_1 up to and including time $t_2 - 1$ and the statistic $r_G(X(t), e)$ increases by at least $\delta/2$ in the time period from t_1 to t_2.

The event that some $r_G(X(\tau), e)$ exceeds $r_{\max}(X(0)) + \delta$ at some time τ, $1 \leq \tau \leq T$ is contained in the event $\bigcup_{e,G} \bigcup_{0 \leq t_1 < t_2 \leq T} B_{t_1,t_2}(e, G, \delta)$. The next claim bounds the probability of the bad event for a particular choice of edge e and graph G and the proof of this lemma follows.

CLAIM 15. *The probability of the event* $\bigcup_{0 \leq t_1 < t_2 \leq T} B_{t_1,t_2}(e, G, \delta)$ *is bounded as*

$$\mathbf{P}\left(\bigcup_{0 \leq t_1 < t_2 \leq T} B_{t_1,t_2}(e, G, \delta)\right) \leq e^{-\Omega(n)}. \tag{12}$$

PROOF. For all X we have $N_{G_i}(X, e, e') \leq N_{G_i}(K_n, e, e')$. The term $N_{G_i}(K_n, e, e')$ is the number of graphs G_i in the complete graph containing both e and e'. In the case that the two edges e and e' share a vertex, they define 3 vertices

which leaves at most $|V_i| - 3$ remaining vertices to be chosen. It follows that $N_{G_i}(K_n, e, e') \le O(n^{|V_i|-3})$ and so

$$\frac{N_{G_i}(X, e, e')}{n^{|V_i|-2}} = O(n^{-1}). \tag{13}$$

Note that an adjacent edge e' is only chosen with probability $O(n^{-1})$. When e and e' do not share an edge, then

$$\frac{N_{G_i}(X, e, e')}{n^{|V_i|-2}} = O(n^{-2}). \tag{14}$$

Although the claim concerns the random variable $r_G(X, e)$, we will work with the related random variable

$$Y_t = \frac{N_G(X(t), e)}{n^{|V|-2}}.$$

The first step is to compute a bound on the moment generating function of

$$S_{t_1,t_2} = \sum_{t=t_1+1}^{t_2} \left(Y_t - Y_{t-1} + \frac{\gamma}{2n^2} \right) \mathbb{1}(D_{t-1}(e, G, \delta)).$$

The random variable S_{t_1,t_2} is the change in Y_i from time t_1 to t_2 while all the edge statistics are within the appropriate interval, shifted by $\frac{\gamma}{2n^2}$ per time step. Clearly, we have the containment

$$B_{t_1,t_2}(e, G, \delta) \subseteq \{S_{t_1,t_2} \ge \delta/2\}. \tag{15}$$

We have

$$\mathbf{E}[e^{\theta S_{t_1,t_2}}] = \mathbf{E}\big[e^{\theta S_{t_1,t_2-1}} \mathbb{E}\big(e^{\theta(Y_{t_2} - Y_{t_2-1} + \gamma/(2n^2))\mathbb{1}(D_{t-1}(e,G,\delta))} | \mathcal{F}_{t_2-1}\big)\big].$$

From Lemma 12 and equation (11) it follows that $\mathbb{E}(Y_t - Y_{t-1}\mathbb{1}(D_{t-1}(e, G, \delta))|\mathcal{F}_{t-1}) \le -\gamma/n^2$. Recalling that with probability $1 - O(n^{-1})$, it holds that $|Y_t - Y_{t-1}| = O(n^{-2})$ and it always holds that $|Y_t - Y_{t-1}| = O(n^{-1})$, we have

$$\begin{aligned}
&\mathbf{E}\big(e^{\theta(Y_{t_2} - Y_{t_2-1} + \gamma/(2n^2))\mathbb{1}(D_{t-1}(e,G,\delta))} | \mathcal{F}_{t_2-1}\big) \\
&\quad = \sum_{k=0}^{\infty} \mathbf{E}\left[\frac{\theta(Y_{t_2} - Y_{t_2-1} + \gamma/(2n^2))^k}{k!} \mathbb{1}(D_{t-1}(e, G, \delta))^k | \mathcal{F}_{t-1} \right] \\
&\quad \le 1 - \mathbb{1}(D_{t-1}(e, G, \delta)) \frac{\gamma\theta}{2n^2} \\
&\qquad + \mathbb{1}(D_{t-1}(e, G, \delta))\theta^2 \\
&\qquad\quad \times \mathbf{E}\left[(Y_t - Y_{t-1})^2 \sum_{k=2}^{\infty} \frac{(\theta(Y_t - Y_{t-1} + \gamma/(2n^2)))^{k-2}}{k!} \Big| \mathcal{F}_{t-1} \right] \\
&\quad = 1 - \mathbb{1}(D_{t-1}(e, G, \delta)) \left(\frac{\gamma\theta}{2n^2} + O\left(\frac{\theta^2}{n^3} \right) \right).
\end{aligned}$$

Thus, when we take $\theta = cn$ for sufficiently small c, we have that

$$\mathbf{E}\big(e^{\theta(Y_{t_2}-Y_{t_2-1}+\gamma/(2n^2))\mathbb{1}(D_{t-1}(e,G,\delta))}|\mathcal{F}_{t-1}\big) \le 1$$

and so

$$\mathbf{E}[e^{\theta S_{t_1,t_2}}] \le \mathbf{E}[e^{\theta S_{t_1,t_2-1}}] \le 1,$$

where the second inequality follows by iterating the argument leading to the first inequality. We can choose $\alpha > 0$ depending only on L and δ such that for any graph in $\{\mathbb{G}_L\}$,

$$\alpha < \sup_{x\in[p^*,1]} \big\{(x+\delta/2)^{|E|-1} - (x)^{|E|-1}\big\}.$$

This gives the estimate

$$\mathbf{P}(S_{t_1,t_2} \ge \alpha) \le e^{-c\alpha n}\mathbb{E}\big[e^{\theta(Y_t-Y_0)}\big] = e^{-\Omega(n)} \tag{16}$$

and so

$$\mathbf{P}\big(r_G(X(t_2),e) - r_G(X(t_1),e) > \delta/2\big) = e^{-\Omega(n)}. \tag{17}$$

We may now apply (17) to equation (15), resulting in

$$\mathbf{P}\Bigg(\bigcup_{0\le t_1\le t\le t_2\le T} B_{t_1,t_2}(e,G,\delta)\Bigg) \le T^2 e^{-\Omega(n)}\big(1+o(n)\big) = e^{-\Omega(n)}, \tag{18}$$

which proves the claim. □

Next, we argue that if all of the random variables $r_G(X(t),e)$ remain below $r_{\max}+\delta$, then each random variable actually ends below $r_{\max}-\delta$ with exponentially high probability. We prove this by showing that each random walk actually reaches $r_{\max}-2\delta$ and then by the claim has exponentially small probability of increasing to $r_{\max}-\delta$. Suppose that for some e, G, $r_G(X(0),e) \ge r_{\max}-2\delta$. Then for $T = cn^2$,

$$\begin{aligned}
&\mathbf{P}\big(r_G(X(t),e) \ge r_{\max}-2\delta \text{ for } 1\le t\le T\big)\\
&\quad\le \mathbf{P}\Bigg(r_G(X(t),e) \ge r_{\max}-2\delta \text{ for } 1\le t\le T, \bigcap_{1\le t\le T} A_t(\delta)\Bigg) + e^{-\Omega(n)}\\
&\quad\le \mathbf{P}\Bigg(r_G(X(t),e) \ge r_{\max}-2\delta \text{ for } 1\le t\le T, \bigcap_{1\le t\le T} D_t(e,G,\delta)\Bigg) + e^{-\Omega(n)}\\
&\quad\le \mathbf{P}\Bigg(S_{1,T} \ge -1+\frac{\gamma c}{2}\Bigg) + e^{-\Omega(n)},
\end{aligned}$$

where the last step follows since each of the T increments in $S_{1,T}$ contribute $\gamma/2n^2$ on the event $\bigcap_{1\le t\le T} D_t(e,G,\delta)$. Choosing $c \ge 3/\gamma$ and using the estimate on the deviation of S_{t_1,t_2}, (17) gives

$$\mathbf{P}\big(r_G(X(t),e) \ge r_{\max}-2\delta \text{ for } 1\le t\le T\big) \le e^{-\Omega(n)}.$$

Finally, we have

$$\begin{aligned}
&\mathbf{P}\big(r_G(X(T),e) \geq r_{\max} - \delta\big) \\
&\quad \leq \mathbf{P}\big(r_G(X(T),e) \geq r_{\max} - \delta, r_G(X(t),e) < r_{\max} - 2\delta \text{ for some } t \in [1,T]\big) \\
&\qquad + e^{-\Omega(n)} \\
&\quad \leq \mathbf{P}\Big(\bigcup_{1 \leq t_1 \leq T} B_{t_1,T}(e,G,\delta)\Big) + e^{-\Omega(n)} \leq e^{-\Omega(n)}.
\end{aligned}$$

The union bound on probabilities applied over the set of edges e and graphs G completes the proof of Lemma 14. □

Our results on $r_{\max}$ can trivially be transferred into the analogous results for $r_{\min}$. The following lemmas follow immediately from iterating Lemma 14.

LEMMA 16. *In the high temperature phase for any $\varepsilon > 0$ there is $c > 0$ such that for any initial configuration $X(0) = x$, when $t \geq cn^2$ we have*

$$\mathbf{P}\big(r_{\max}(X(t)) \geq p^* + \varepsilon | X(0) = x\big) \leq e^{-\Omega(n)},$$
$$\mathbf{P}\big(r_{\min}(X(t)) \leq p^* - \varepsilon | X(0) = x\big) \leq e^{-\Omega(n)}.$$

LEMMA 17. *In the low temperature phase suppose that p^* is a solution to $p = \varphi(p)$ and $\varphi'(p^*) < 1$. There exists an $\epsilon > 0$ such that if for some initial configuration $X(0)$, we have that $r_{\max}(X(0)) \leq p^* + \varepsilon$ and $r_{\min}(X(0)) \geq p^* - \varepsilon$ then for some $\alpha > 0$*

$$\mathbf{P}\Big(\sup_{0<t<e^{\alpha n}} r_{\max}(X(t)) \geq p^* + 2\varepsilon\Big) \leq e^{-\Omega(n)},$$
$$\mathbf{P}\Big(\inf_{0<t<e^{\alpha n}} r_{\min}(X(t)) \leq p^* - 2\varepsilon\Big) \leq e^{-\Omega(n)}.$$

2.3. *Path coupling.*

LEMMA 18. *Let $p^* \in [0,1]$ be a solution of the equation $\varphi(p) = p$ and suppose $0 < \varphi'(p^*) < 1$. There exists $\epsilon, \delta > 0$ sufficiently small and such that the following holds. Suppose that $X^+(0) \geq X^-(0)$ are two configurations that differ at exactly one edge e. Suppose further that for all graphs G with at most L vertices and all edges e'*

$$|r(G,e') - p^*| < \epsilon. \tag{19}$$

Then for sufficiently large n, a single step of the Glauber dynamics can be coupled so that

$$\mathbf{E}d_H(X^+(1), X^-(1)) \leq 1 - \delta n^{-2}.$$

PROOF. We take the standard monotone coupling. Suppose that an edge $e' \neq e$ is chosen to be updated by the Markov chain. Then

$$\mathbf{P}\big(X^{\pm}_{e'}(1) = 1\big) = \frac{\exp(\partial_{e'} H(X^{\pm}(0)))}{1 + \exp(\partial_{e'} H(X^{\pm}(0)))}. \tag{20}$$

Since

$$\partial_{e'} H(X^{\pm}(0)) = \sum_{i=1}^{s} \frac{\beta_i N_{G_i}(X^{\pm}(0), e')}{n^{|V_i|-2}}$$

by Lemma 9 and equation (19), we have that for large enough n,

$$\partial_{e'} H(X^{\pm}(0)) \leq \sum_{i=1}^{s} \frac{\beta_i (p^* + \epsilon)^{|E_i|-1} N_{G_i}(K_n, e')}{n^{|V_i|-2}} = \Psi_\beta(p^* + \epsilon). \tag{21}$$

Similarly,

$$0 \leq \big(1 - o(1)\big)\Psi(p^* - \epsilon) \leq \partial_{e'} H(X^{\pm}(0))$$

and so it follows that for any $\epsilon' > 0$, for large enough n and for small enough ϵ we have that

$$\frac{d}{dx}\frac{e^x}{1+e^x}\bigg|_{\partial_{e'} H(X^{+}(0))} \leq (1+\epsilon')\,\frac{d}{dx}\frac{e^x}{1+e^x}\bigg|_{\Psi(p^*)}. \tag{22}$$

We now bound the sum of the $\partial_e\,\partial_{e'} H(X^{+}(0))$ terms

$$\begin{aligned}
\sum_{e' \neq e} \partial_e\,\partial_{e'} H(X^{+}(0)) &= \sum_{e' \neq e} \sum_{i=1}^{s} \frac{\beta_i N_{G_i}(X^{+}(0), e, e')}{n^{|V_i|-2}} \\
&= \sum_{i=1}^{s} \frac{\beta_i \sum_{\substack{\alpha \in E(G_i) \\ \alpha \neq e}} N_{(G_i)_\alpha}(X^{+}(0), e)}{n^{|V_i|-2}} \\
&\leq \sum_{i=1}^{s} \frac{\beta_i \sum_{\substack{\alpha \in E(G_i) \\ \alpha \neq e}} (p^* + \epsilon)^{|E_i|-2} N_{(G_i)_\alpha}(K_n, e)}{n^{|V_i|-2}} \\
&= \sum_{i=1}^{s} \sum_{e' \neq e} \frac{\beta_i (p^* + \epsilon)^{|E_i|-2} N_{G_i}(K_n, e, e')}{n^{|V_i|-2}},
\end{aligned}$$

where the second and fourth lines follow from Lemma 10 and the inequality follows from equation (19). By Lemma 9 we have that

$$\begin{aligned}
\sum_{e' \neq e} \partial_e\,\partial_{e'} H(X^{+}(0)) &\leq \big(1 + o(1)\big) \sum_{i=1}^{s} 2|E_i|(|E_i| - 1)(p^* + \epsilon)^{|E_i|-2} \\
&= \big(1 + o(1)\big)\Psi'(p^* + \epsilon).
\end{aligned} \tag{23}$$

By the Taylor series for small h we have that

$$\frac{e^{x+h}}{1+e^{x+h}} - \frac{e^x}{1+e^x} \le \frac{d}{dx}\frac{e^x}{1+e^x}\bigg|_x (h + O(h^2))$$

and so using equation (20),

$$\begin{aligned}
&\mathbf{P}\big(X^+_{e'}(1)=1\big) - \mathbf{P}\big(X^-_{e'}(1)=1\big) \\
&\quad = \frac{d}{dx}\frac{e^x}{1+e^x}\bigg|_{\partial_{e'}H(X^+(0))} \cdot \big(\partial_e\,\partial_{e'}H(X^+(0)) + O(\partial_e\,\partial_{e'}H(X^+(0))^2)\big) \qquad (24)\\
&\quad \le (1+\epsilon')\big(1+o(1)\big)\,\partial_e\,\partial_{e'}H(X^+(0))\,\frac{d}{dx}\frac{e^x}{1+e^x}\bigg|_{\Psi(p^*)},
\end{aligned}$$

using equations (22) and the fact that by equation (13) we have that

$$\partial_e\,\partial_{e'}H(X^+(0)) = O(n^{-1}).$$

Each edge e' has probability $\binom{n}{2}^{-1}$ of being updated and if edge e is chosen to be updated then the number of disagreements is 0. It follows by equations (24) and (23) that for any $\epsilon'' > 0$

$$\begin{aligned}
&\mathbf{E}d_H(X^+(1), X^-(1)) \\
&\quad \le 1 - \binom{n}{2}^{-1}\Bigg[1 - \sum_{e'\neq e}(1+\epsilon')\big(1+o(1)\big)\,\partial_e\,\partial_{e'}H(X^+(0))\,\frac{d}{dx}\frac{e^x}{1+e^x}\bigg|_{\Psi(p)}\Bigg] \\
&\quad \le 1 - \binom{n}{2}^{-1}\Bigg[1 - (1+\epsilon')\big(1+o(1)\big)\Psi'(p^*+\epsilon)\,\frac{d}{dx}\frac{e^x}{1+e^x}\bigg|_{\Psi(p^*)}\Bigg] \\
&\quad \le 1 - \binom{n}{2}^{-1}\big[1 - (1+\epsilon'')\big(1+o(1)\big)\varphi'(p^*)\big]
\end{aligned}$$

provided that ϵ, ϵ' are sufficiently small. The result follows, since $\varphi'(p^*) < 1$. □

PROOF OF THEOREM 5. We begin by proving the high temperature phase using a coupling argument. Let $X^+(t)$ and $X^-(t)$ be two copies of the Markov chain started from the complete and empty configurations, respectively, and coupled using the monotone coupling. Since this is a monotone system, it follows that if $\mathbf{P}(X^+(t) \neq X^-(t)) < e^{-1}$, then t is an upper bound on the mixing time. The function φ satisfies the hypothesis of Lemma 18, so choose ϵ and δ according to the lemma. Let property $\mathcal{A}_t$ be the event that for all graphs $G \in \mathbb{G}_L$ and all edges e

$$|r_G(X, e) - p^*| < \epsilon \qquad (25)$$

for both $X^+(t)$ and $X^-(t)$. By Lemma 16 we have that if $t \ge cn^2$, then $\mathbf{P}(\mathcal{A}_t) \ge 1 - e^{-\alpha n}$.

Since $X^+(t) \geq X^-(t)$, there exists a sequence of configurations $X^-(t) = X^0 \leq X^1 \leq \cdots \leq X^d = X^+(t)$, where $d = d_H(X^+(t), X^-(t))$, each pair X^i, X^{i+1} differ at exactly one edge and each X^i satisfies equation (25). Such a sequence is constructed by adding one edge at a time to $X^-(t)$ until $X^+(t)$ is reached. Further, since the subgraph counts $N_G(X, e)$ are monotone in X, if both $X^+(t)$ and $X^-(t)$ satisfy (19), each of the configurations X^i also satisfies this equation. Applying path coupling to this sequence, we have that by Lemma 18

$$\begin{aligned}&\mathbf{E}\big[d_H\big(X^+(t+1), X^-(t+1)\big)\mathbb{1}(\mathcal{A}_t)|X^+(t), X^-(t)\big]\\&\quad\leq (1-\delta n^{-2})d_H(X^+(t), X^-(t)).\end{aligned}$$

Since $d_H(X^+(t), X^-(t)) \leq \binom{n}{2}$, we have the inequality

$$\begin{aligned}&\mathbf{E}\big[d_H\big(X^+(t+1), X^-(t+1)\big)\big]\\&\quad\leq (1-\delta n^{-2})\mathbf{E}[d_H(X^+(t), X^-(t))] + \binom{n}{2}\big(1-P(\mathcal{A}_t)\big)\\&\quad\leq (1-\delta n^{-2})\mathbf{E}[d_H(X^+(t), X^-(t))] + \binom{n}{2}e^{-\alpha n}.\end{aligned}$$

Iterating this equation, we get that for $t > Cn^2$,

$$\begin{aligned}&\mathbf{E}[d_H(X^+(t), X^-(t))]\\&\quad\leq (1-\delta n^{-2})^{t-Cn^2}\binom{n}{2} + e^{-\alpha n}\binom{n}{2}\sum_{j=C'n^2}^{t}(1-\delta n^{-2})^{t-j}\\&\quad\leq \exp\big(-\delta n^{-2}(t-Cn^2)\big)n^2 + e^{-\alpha n}\frac{1}{\delta}\binom{n}{2}n^2.\end{aligned}$$

Then for any $\epsilon' > 0$, when $t > \frac{C+2+\epsilon'}{\delta}n^2\log n$, we have that for large enough n,

$$\mathbf{E}[d_H(X^+(t), X^-(t))] = o(1).$$

It follows by Markov's inequality that $\mathbf{P}(X^+(t) \neq X^-(t)) = o(1)$, which establishes that the mixing time is bounded by $\frac{C+2+\epsilon'}{\delta}n^2\log n$.

To establish a lower bound of order $n^2\log n$, we note that at time $t^\star := \frac{1}{2}n^2\log n$ by a coupon collecting argument there are of order $n^{3/2}$ edges which have not yet been updated. Now let $\mathcal{GAP}$ denote the spectral gap of the Glauber dynamics, that is, the difference of the largest and second largest eigenvalues by absolute value. It is well known that for a Markov chain with state space Ω, transition matrix P and stationary distribution π that

$$\mathcal{GAP} = \inf_{f:Ef\neq 0}\frac{\sum_{s,s'\in S}(f(s)-f(s'))^2P(s,s')\pi(s)}{\operatorname{var} f}$$

and that $\mathcal{GAP}^{-1} \le \tau_{\mathrm{mix}}$ (see, e.g., [2]). Let X be chosen according to the stationary distribution and take $Y = Y(X)$ to denote the number of edges in the graph. Since at most one edge is changed in each step of the Markov chain we have that

$$\operatorname{var} Y \le \mathcal{GAP}^{-1} \le \tau_{\mathrm{mix}} = O(n^2 \log n).$$

[Actually, it can be shown that $\mathcal{GAP}^{-1} = O(n^2)$ but we do not need it for this argument.] By Chebyshev's inequality we have that

$$P(|Y - \mathbf{E}Y| \le n^{4/3}) = O(n^{-2/3} \log n).$$

On the other hand, we know that with high probability $Y(X^+(t^\star)) - Y(X^-(t^\star)) > cn^{3/2}$, so at most one of $Y(X^+(t^\star))$ or $Y(X^-(t^\star))$ can be close to $\mathbf{E}Y$. Hence, we have that

$$\max\{\mathbf{P}(|Y(X^+(t^\star)) - \mathbf{E}Y| \le n^{4/3}), \mathbf{P}(|Y(X^-(t^\star)) - \mathbf{E}Y| \le n^{4/3})\} \ge \tfrac{1}{2} - o(1).$$

This, in particular, means that either $Y(X^+(t^\star))$ or $Y(X^-(t^\star))$ (and actually it can be shown that both) is not within e^{-1} in total variation distance from the stationary distribution. Hence, the mixing time is at least $t^\star = \frac{1}{2}n^2 \log n$ when n is large which completes the proof. □

2.4. *Slow mixing for local Markov chains in low temperature regime.* We will use the following conductance result, which is taken from [9], Claim 2.3:

CLAIM 19. *Let $\mathcal{M}$ be a Markov chain with state space Ω, transition matrix P and stationary distribution π. Let $A \subset \Omega$ be a set of states such that $\pi(A) \le \frac{1}{2}$ and $B \subset \Omega$ be a set of states that form a "barrier" in the sense that $P_{ij} = 0$ whenever $i \in A \setminus B$ and $j \in A^c \setminus B$. Then the mixing time of $\mathcal{M}$ is at least $\pi(A)/8\pi(B)$.*

Using this result we prove slow mixing for any local Markov chain.

PROOF OF THEOREM 6. Suppose p_1 and p_2 are solutions of the equation $\varphi(p) = p$ with $\varphi'(p_1) < 1, \varphi'(p_2) < 1$ and choose $\varepsilon > 0$ sufficiently small so that $\varphi(p) < p$ for $p \in (p_i, p_i + 3\varepsilon]$ and $\varphi(p) > p$ for $p \in [p_i - 3\varepsilon, p_i)$, for $i = 1, 2$. Let

$$A_i = \{X : r_{\max}(X) \le p_i + \varepsilon \text{ and } r_{\min}(X) \ge p_i - \varepsilon\}, \qquad i = 1, 2$$

and suppose the set A_1 has smaller probability (switching the labels p_1 and p_2 if necessary), so $\pi(A_1) \le \frac{1}{2}$. We note that for large enough n, $\pi(A_i) > 0$ since with high probability an Erdős–Rényi random graph $G(n, p_i)$ is in A_i. In the remainder of the proof we will omit the subscript, that is, let $A = A_1$ and $p = p_i$. Now, clearly the set

$$B = \{X : p + \varepsilon < r_{\max}(X) \le p + 2\varepsilon \text{ or } p - \varepsilon > r_{\min}(X) \ge p - 2\varepsilon\}$$

forms a barrier (for sufficiently large n) between the sets A and A^c for any Markov chain that updates only $o(n)$ edges per time-step, since each edge update can change each of $r_{\max}$ and $r_{\min}$ by at most $O(\frac{1}{n})$.

It remains only to bound the relative probabilities of the sets A and B. Let $C = A^c \setminus B$ and let $t = cn^2$ such that Lemma 14 holds. Then

$$\mathbf{P}\big(X(t) \in C | X(0) \in B\big) = e^{-\Omega(n)} \tag{26}$$

and

$$\mathbf{P}\big(X(t) \in B | X(0) \in A \cup B\big) = e^{-\Omega(n)}. \tag{27}$$

Let the configuration $X(0)$ be drawn according to the Gibbs measure $\pi = p_n$ defined in equation (3) and let $X(t)$ be the configuration resulting after t steps of the Glauber dynamics. Because the Glauber dynamics has stationary distribution π, $X(t)$ has the same distribution as $X(0)$. By the reversibility of the Glauber dynamics and the estimate (26), we have

$$\begin{aligned} \mathbf{P}\big(X(t) \in B, X(0) \in C\big) &= \mathbf{P}\big(X(t) \in C, X(0) \in B\big) \\ &= \mathbf{P}\big(X(t) \in C | X(0) \in B\big) P\big(X(0) \in B\big) \\ &= e^{-\Omega(n)} \mathbf{P}\big(X(0) \in B\big). \end{aligned} \tag{28}$$

Similarly, using (27),

$$\begin{aligned} &\mathbf{P}\big(X(t) \in B, X(0) \in A \cup B\big) \\ &\quad = \mathbf{P}\big(X(t) \in B | X(0) \in A \cup B\big) \mathbf{P}\big(X(0) \in A \cup B\big) \\ &\quad \le e^{-\Omega(n)} \mathbf{P}\big(X(0) \in A \cup B\big) \\ &\quad = e^{-\Omega(n)} \big(\mathbf{P}\big(X(0) \in A\big) + \mathbf{P}\big(X(0) \in B\big)\big). \end{aligned} \tag{29}$$

Combining (28) and (29), we have

$$\begin{aligned} \pi(B) &= \mathbf{P}\big(X(t) \in B\big) \\ &= \mathbf{P}\big(X(t) \in B, X(0) \in C\big) + \mathbf{P}\big(X(t) \in B, X(0) \in A \cup B\big) \\ &\le e^{-\Omega(n)} \big(\mathbf{P}\big(X(0) \in A\big) + 2\mathbf{P}\big(X(0) \in B\big)\big) \\ &= e^{-\Omega(n)} \big(\pi(A) + 2\pi(B)\big), \end{aligned} \tag{30}$$

which, upon rearranging, gives

$$\pi(B) \le \frac{e^{-\Omega(n)}}{1 - 2e^{-\Omega(n)}} \pi(A). \tag{31}$$

Applying Claim 19 completes the proof. $\square$

3. Independence of edges and weak pseudo-randomness. Here we shall prove Theorem 7. Note that our burn-in proof in the high temperature regime shows that with high probability all the $r_G(X, e)$ are close to p^*, the fixed point of $\varphi(p) = p$. A consequence is that for any collection of edges $e_1, \dots, e_j$ the events x_{e_i} are asymptotically independent and distributed as Bernoulli(p^*). A consequence of the asymptotic independence of the edges is that with high probability a graph sample from the exponential random graph distribution is *weakly pseudo-random*, as defined in [14]. As such, the exponential random graph model (in the high temperature case) is similar to the basic Erdős–Rényi random graph. Since exponential random graphs were introduced to model the phenomenon of increased numbers of small subgraphs like triangles, this result suggests that in the high temperature case, this model does not quite achieve its goal.

PROOF OF THEOREM 7. Fix $\varepsilon > 0$. Let $S \subseteq [k]$ and let $x_S = \{x_{e_i} : i \in S\}$ and $x_{S^c} = \{x_{e_i} : i \in [k] \setminus S\}$. Then, by the inclusion-exclusion principle, we have

$$\mathbf{P}(x_S = 1, x_{S^c} = 0) = \sum_{T \subseteq [k]:\, S \subseteq T} (-1)^{|T|-|S|} \mathbf{P}(x_T = 1). \tag{32}$$

Hence, it suffices to show that asymptotically each probability in the preceding sum satisfies $\mathbf{P}(x_T = 1) = (1 + o(1))(p^*)^{|T|}$. Define

$$A = \{X : r_{\max}(X) \le p^* + \varepsilon \text{ and } r_{\min}(X) \ge p^* - \varepsilon\}.$$

Consider the subgraph G_T formed by the edges in the set T. Conditional on $X \in A$, $|r_{G_T}(X, e) - p^*| \le \varepsilon$, which gives

$$\left|N_{G_T}(X, e) - (p^*)^{|E|-1} 2|E| n^{|V|-2}\right| = O(\varepsilon), \tag{33}$$

where V is the number of vertices and $|E| = |T|$ is the number of edges of G_T. By considering the graph consisting of two disjoint edges, we have that the number of edges in a configuration $X \in A$ satisfies

$$\left|N_{\text{edge}}(X) - p^* \binom{n}{2}\right| = O(\varepsilon). \tag{34}$$

Note that

$$\sum_{e \in X} N_G(X, e) \mathbb{1}(e_i \in X) = |E| N_G(X)$$

and summing equation (33) over the edges in X and using (34) gives

$$\left|N_{G_T}(X) - (p^*)^{|E|} n^{|V|}\right| \le O(\varepsilon) \tag{35}$$

for sufficiently large n. By symmetry, each of the subgraphs G_T is equally likely to be included in the configuration X and there are $n^{|V|}$ possible such subgraphs, so

$$\mathbf{P}(x_T = 1 | X \in A) = \frac{N_{G_T}(X)}{n^{|V|}} = (p^*)^{|E|} + O(\varepsilon).$$

Recall that $\mathbf{P}(A) = 1 - o(1)$ in the high temperature phase (as argued in the proof of Theorem 5 using Lemma 16). Thus, for any set of edges $T \subseteq [k]$, we have that

$\mathbf{P}(x_T = 1) = (p^*)^{|T|} + O(\varepsilon)$ where we used the fact that $|E| = |T|$. Plugging this into (32) and using the fact that

$$\sum_{T\subseteq[k]:S\subseteq T} (-1)^{|T|}(p^*)^{|T|} = (p^*)^{|S|} \sum_{q=0}^{k-|S|} \binom{k-|S|}{q} (-p^*)^q$$
$$= (p^*)^{|S|}(1-p^*)^{k-|S|},$$

we have that for large n

$$\left|\mathbf{P}(x_1 = a_1, \ldots, x_k = a_k) - (p^*)^{\sum a_i}(1-p^*)^{k-\sum a_i}\right| \le O(\varepsilon),$$

which completes the proof by taking ε to 0. □

We can also show that an exponential random graph is weakly pseudo-random with high probability. This means that a collection of equivalent conditions are satisfied; we briefly mention only a few of them (see the survey on pseudo-random graphs [14]). We will use a different subgraph count than before: for a graph G let $N_G^*(X)$ be the number of labeled induced copies of G in X. This is different than the counts $N_G(X)$ in that it requires edges missing from G to also be missing in the induced graph in X. A graph X is weakly pseudo-random if it satisfies one of the following (among others) equivalent properties:

1. For a fixed $l \ge 4$ for all graphs G on l vertices,
$$N_G^*(X) = \big(1+o(1)\big)n^l(p)^{|E(G)|}(1-p)^{\binom{l}{2}-|E(G)|}.$$
2. $N_{\text{edges}}(X) \ge \frac{n^2p}{2} + o(n^2)$ and $\lambda_1 = (1+o(1))np$, $\lambda_2 = o(n)$, where the eigenvalues of the adjacency matrix of X are ordered so that $|\lambda_1| \ge |\lambda_2| \ge \cdots \ge |\lambda_n|$.
3. For each subset of vertices $U \subset V(X)$ the number of edges in the subgraph of X induced by U satisfies $E(H_X(U)) = \frac{p}{2}|U|^2 + o(n^2)$.
4. Let C_l denote the cycle of length l, with $l \ge 4$ even. The number of edges in X satisfies $N_{\text{edges}}(X) = \frac{n^2p}{2} + o(n^2)$ and the number of cycles C_l satisfies $N_{C_l}(X) \le (np)^l + o(n^l)$.

By Theorem 7, for any configuration in the good set, $X \in \mathbf{G}$, the fourth condition is satisfied. This gives the following corollary.

COROLLARY 20 (Weak pseudo-randomness). *With probability* $1 - o(1)$, *an exponential random graph is weakly pseudo-random.*

Acknowledgments. We gratefully acknowledge the detailed and thorough comments from the anonymous reviewers.

REFERENCES

[1] ALBERT, R. and BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* **74** 47–97. MR1895096

[2] ALDOUS, D. and FILL, J. (2011). Reversible Markov chains and random walks on graphs. Unpublished manuscript.

[3] ANDERSON, C. J., WASSERMAN, S. and CROUCH, B. (1999). A p^* primer: Logit models for social networks. *Social Networks* **21** 37–66.
[4] BUBLEY, R. and DYER, M. (1997). Path coupling: A technique for proving rapid mixing in Markov chains. In *Foundations of Computer Science* (*FOCS*) 223–231. IEEE.
[5] CHATTERJEE, S. and VARADHAN, S. R. S. (2011). The large deviation principle for the Erdos–Renyi random graph. *European J. Comb.* To appear. Available at arXiv:1008.1946v2.
[6] CHUNG, F. R. K., GRAHAM, R. L. and WILSON, R. M. (1989). Quasi-random graphs. *Combinatorica* **9** 345–362. MR1054011
[7] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge Univ. Press, Cambridge. MR2271734
[8] DYER, M. and FRIEZE, A. (2003). Randomly coloring graphs with lower bounds on girth and maximum degree. *Random Structures Algorithms* **23** 167–179. MR1995689
[9] DYER, M., FRIEZE, A. and JERRUM, M. (2002). On counting independent sets in sparse graphs. *SIAM J. Comput.* **31** 1527–1541 (electronic). MR1936657
[10] FRANK, O. and STRAUSS, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. MR0860518
[11] JAYNES, E. T. (1957). Information theory and statistical mechanics. *Phys. Rev.* (2) **106** 620–630. MR0087305
[12] JAYNES, E. T. (1957). Information theory and statistical mechanics. II. *Phys. Rev.* (2) **108** 171–190. MR0096414
[13] KASS, R. E. and WASSERMAN, L. (1996). The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.* **91** 1343–1370.
[14] KRIVELEVICH, M. and SUDAKOV, B. (2005). Pseudo-random graphs. In *Conference on Finite and Infinite Sets* 1–64. Bolyai Mathematical Society, Budapest.
[15] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45** 167–256 (electronic). MR2010377
[16] PARK, J. and NEWMAN, M. E. J. (2004). Solution of the 2-star model of a network. *Phys. Rev. E* (3) **70** 066146.1–066146.5.
[17] PARK, J. and NEWMAN, M. E. J. (2005). Solution for the properties of a clustered network. *Phys. Rev. E* (3) **72** 026136.1–026136.5.
[18] SNIJDERS, T. A., PATTISON, P., ROBBINS, G. and HANDCOCK, M. (2006). New specifications for exponential random graph models. *Sociological Methodology* **36** 99–153.
[19] WASSERMAN, S. and PATTISON, P. (1996). Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and p. *Psychometrika* **61** 401–425. MR1424909

A.BHAMIDI
DEPARTMENT OF STATISTICS
AND OPERATIONS RESEARCH
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599
USA
E-MAIL: bhamidi@email.unc.edu

G. BRESLER
DEPARTMENT OF EECS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: gbresler@eecs.berkeley.edu

A. SLY
MICROSOFT RESEARCH
ONE MICROSOFT WAY
REDMOND, WASHINGTON 98052
USA
E-MAIL: allansly@microsoft.com