Mixing time of exponential random graphs

Shankar Bhamidi* Guy Bresler† Allan Sly‡

April 10, 2008

Abstract

A plethora of random graph models have 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 the 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 answering almost any problem of parameter estimation hypothesis testing or to understand the inherent network model itself. In practice this sampling is typically carried out using either the Glauber dynamics or the Metropolis-Hasting Markov chain Monte Carlo procedure.

In this paper we characterize the high and low temperature regimes. 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 MCMC is possible, we show that any finite collection of edges are 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, largely driven by the availability of new data as well as the growth of computing resources, there has been explosion in the study of various real-world networks such as 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 for example [10, 11, 12]).

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 friendship. The basic idea is that if two individuals share a common friend, then they are more likely than

*Department of Statistics, University of California, Berkeley. Email: shanky@stat.berkeley.edu. Research supported by N.S.F. Grant DMS 0704159

†Department of Electrical Engineering and Computer Sciences, University of California, Berkeley. Email: gbresler@eecs.berkeley.edu. Research supported by Vodafone US-Foundation fellowship

‡Department of Statistics, University of California, Berkeley. Email: sly@stat.berkeley.edu. Research supported by N.S.F. Grants DMS 0528488 and DMS 0548249

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. 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

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

Then a probability measure on the space of graphs is 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,

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

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, and 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 in general 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 [6]. Being able to sample from exponential random graphs to observe their large scale properties is one way this can be realised. Sampling is almost always carried out using local MCMC algorithms, particularly the Glauber dynamics or Metropolis-Hasting. 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 to the Glauber dynamics, one of the most popular Markov chains in this context. We provide the first rigorous analysis of the mixing time of the Glauber dynamics for sampling this distribution. In the process we also provide 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 Sourav Chatterjee in the case of triangles) we can nonetheless show that the distribution has poor conductance and thus establish exponentially slow mixing for any local Markov chain with the specified stationary distribution.

Relevant 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. The pioneering article in this area by Frank and Strauss [9] introduced the concept of Markov graphs. They were among the first to introduce the notion of exponential random graphs but in the limited situation where the subgraphs are stars or triangles. Extending the methodology of [9], general subgraph counts, were introduced in [8]. However, from the outset a number of researchers noted problems at the empirical level for their Markov chain algorithms, depending on parameter values. See [6] for a relevant discussion of empirical findings as well as many new specifications of the model to circumvent such issues. The interested reader is refered to the comprehensive survey length articles [6] and [5] for a description of the contexts in which these models arise and the statistical methodologies that have been developed around them as well as an extensive bibliography; also see the more general article [10] which surveys a wide variety of common random graph models.

Sourav Chatterjee [4], in his recent work characterizing the large deviation properties of Erdős-Rényi random graphs, developed mathematical techniques which can be used to study the actual structure of these random graphs. At the statistical physics level Mark Newman and his co-authors have made fundamental contributions in the case where the subgraphs are triangles and 2-stars see [2] and [3].

1.1 Definitions and Notation

This section contains a precise mathematical definition of the model and the Markov chain methodology we use. We work on the space $\mathcal{G}_n$ of all graphs on n vertices with vertex set $[n] := \{1, 2, \ldots, n\}$. We shall use $X = (x_e)$ to denote a graph from $\mathcal{G}_n$ where for every 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. 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:

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

We shall denote such an m tuple of distinct vertices by $\mathbf{v}_m$. In graph X, for any m distinct vertices $\mathbf{v}_m$, let $H_X(\mathbf{v}_m)$ denote the subgraph induced by X on $\mathbf{v}_m$. 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 the configuration $X \in \mathcal{G}_n$ and the 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 constants. 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 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:

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

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 $k \geq 1$ and fix graphs $G_1, G_2 \ldots, G_s$ with G_i a graph on $|V_i|$ labeled vertices, with $|V_i| \leq L$ 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 $i \geq 2$ and the $\beta_1 \in \mathbb{R}$.

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

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

Here $Z_n(\beta)$ is the normalizing factor and is often called the partition function and for simplicity we have suppressed the dependence of the measure on the vector β. Also note the normalizations of the subgraph counts of G_i by the factor $n^{|V|-2}$ so that contribution of each factor scales properly and is of order n^2 in the large n limit. Setting $\beta_i \geq 0$ for $i \geq 2$ makes the Gibbs measure a monotone (also ferromagnetic) system which will be important for our proof. The term β_1 does not effect the interaction between edges and plays the role of an external field in this model, making it more or less likely for edges to be included. This is the measure we shall be interested in sampling from. The term in the exponent is often called the *Hamiltonian* and we shall denote it by:

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

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

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

Here A_e consists of all terms dependent on edge e and $B_e(X)$ denotes all terms independent of edge e. Writing X_{e+} as the configuration of edges which coincides with X for edges $e \neq f$ and $X_{e+}(e) = 1$ we define the partial derivative with respect to the edge e of the Hamiltonian H, evaluated at a configuration X 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.

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 above, the corresponding Glauber dynamics is a discrete time ergodic Markov chain on $\mathcal{G}_n$ where given the current state X, the next state X' is obtained by choosing an edge 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 sample from the above Gibbs measure. Call a chain on $\mathcal{G}_n$ *local* if there is a function $\omega_n = o(n)$ which updates at most ω_n edges 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 $\frac{\exp(\partial_e H(X))}{1+\exp(\partial_e H(X))}$ while 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 fundamental lemma is well know, see e.g. [1].

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*

$$\max x, y[\mathbb{P}(X_{t_0} \neq Y_{t_0} | X_0 = x, Y_0 = y) \leq (2e)^{-1}.$$

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

Because 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, $X(t) \leq Y(t)$, where 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 $\beta \in \mathbb{R} \times (\mathbb{R}_+)^{s-1}$ the functions

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

$$\varphi_\beta(p) = \frac{\exp(\Psi(p))}{1+\exp(\Psi(p))}$$

Note that Ψ_β 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. 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 $\varphi_\beta(p) = p$.

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 high temperature phase if $\varphi_\beta(p^*) = p^*$ has a unique fixed point p^* which satisfies $\varphi'_\beta(p^*) < 1$.

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. For simplicity, in the proof we shall suppress the dependence of the functions on β and write φ for φ_β and Ψ for Ψ_β.

1.2 Results

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 high 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 $\mathcal{G}_n$.*

1.3 Idea of the proof

We now give an overview of the proof:

- 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* [15], where 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

  $$\mathbb{E}(d_H(X(1),Y(1))|X(0) = x, Y(0) = y) \leq (1-\beta) \tag{5}$$

  for some $\beta = \beta(n)$ where d_H is the Hamming distance. However this approach fails for some φ_β in the high temperature regime when $\sup_{0\leq p\leq 1} \varphi'(p) > 1$.

- 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 $\varphi'_\beta(\cdot) < 1$), in which path coupling does give a contraction. More precisely, for a configuration X, define

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

  This is (asymptotically) the maximum likelihood choice for the parameter p of the Erdös-Renyi random graph on n vertices, $G(n, p)$, having observed $N_G(X, e)$ subgraphs G containing the edge e. Let $\{G_\lambda\}$ denote the class of all graphs with at most L vertices where L is some integer greater than or equal to $\max_i |V_i|$. Then what we prove is that for ε small enough if the two configurations x and y belong to the set:

$$\mathbf{G} := \{X : \max_{G \in G_\lambda} |r_G(X, e) - p^*| < \varepsilon\}$$

  then equation 5 holds for $\beta(n) = -\delta/n^2$ 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 [16].

- To show that we get into the good set $\mathbf{G}$ quickly, we control all the $r_G(X, e)$, for all subgraphs $G \in G_\lambda$ 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 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 Erdös-Renyi $G(n, p^*)$ random graph model.

- In the low temperature phase, we use a conductance argument developed in [14] 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 Erdös-Renyí 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, thus 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 Erdös-Renyí random graph with parameter p^*. Thus mixing takes an exponentially long time.

2 Proof of the main results

For a fixed graph G and graph $X \in \mathcal{G}_n$, recall the subgraph counts $N_G(X)$, $N_G(X, e)$, and $N_G(X, e, e')$ as defined in Section 1.1. The following lemma follows easily from the definitions,

and its proof is omitted.

Lemma 7 *For an edge α in the graph G, denote by G_α the graph obtained from G by removing the edge α. Then*

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

2.1 Burn-in period

Recall the definition of $r_G(X, e)$ (see Equation (6)):

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

and let $r_{\max}(X) = \max_{e,\lambda} r_{G_\lambda}(X, e)$.

Lemma 8 *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 $p^* + \mu \leq r_{max}(X(0)) \leq \bar{p} - \mu$ for some $\mu > 0$. Then there is a $\delta, c > 0$, depending only on μ, L and φ, so that after $T = c_1 n^2$ steps of the Glauber dynamics, it holds that $r_{max}(X(T)) \leq r_{max}(X(0)) - \delta$ with probability $1 - e^{-\Omega(n)}$.*

Proof. We use the following claim, which shows a negative drift for the random variables $r_G(X, e)$ in the neighborhood of $r_{\max}(X)$.

Claim 9 *The expected change in $N_G(X, e)$ after one step of the Glauber dynamics, starting from the configuration X, can be bounded as*

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

The proof is completed essentially by coupling each of the $r_G(X, e)$ to biased random walks, and is deferred to the Appendix.

The following lemmas follow immediately from iterating Lemma 8 a finite number of times and using the union bound on probabilities.

Lemma 10 *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*

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

Lemma 11 *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 $\alpha > 0$*

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

2.2 Path coupling

Lemma 12 *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'*

$$\left|r(G,e') - p^*\right| < \epsilon. \tag{8}$$

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

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

Proof. The proof is deferred to the Appendix.

Proof of Theorem 5

We begin by proving the high temperature phase using a coupling argument. Let $X^+(t)$ and $X^-(0)$ 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 $P(X^+(t) \neq X^-(t)) < e^{-1}$ then t is an upper bound on the mixing time. The function φ satisfies the hypothesis of Lemma 12 so choose ϵ and δ according to the lemma. Let property $\mathcal{A}_t$ be the event that for all graphs G with at most L vertices and all edges e

$$|r(G,e) - p^*| < \epsilon. \tag{9}$$

for both $X^+(t)$ and $X^-(0)$. By Lemma 10 we have that when $t \geq cn^2$ then $P(\mathcal{A}_t) \geq 1 - e^{-\alpha n}$.

Since the $N_G(X,e)$ are monotone in X if $X^+(t)$ and $X^-(t)$ both satisfy equation (9) then there exists a sequence of configurations $X^-(t) = X^0 \leq X^1 \leq \ldots \leq X^d = X^+(t)$ where $d = d_H(X^+(t), X^-(t))$ and each pair X^i, X^{i+1} differ at exactly one edge and each X^i satisfies equation (9). Such a sequence is constructed by adding one edge at a time to $X^-(t)$ until $X^+(t)$ is reached. Applying path coupling to this sequence we have that by Lemma 12

$$E\left[d_H(X^+(t+1), X^-(t+1))|X^+(t), X^-(t+1), \mathcal{A}_t\right] \leq (1 - \delta n^{-2}) d_H(X^+(t), X^-(t)).$$

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

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

Iterating this equation for we get that $t > C'n^2$,

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

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

$$E\left[d_H(X^+(t), X^-(t))\right] = o(1)$$

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

2.3 Slow mixing for local Markov chains in low-temperature regime

We will use the following conductance result, which is taken from [14] (Claim 2.3):

Claim 13 *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 $\pi(A) \leq \frac{1}{2}$, 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 $\pi(A)/8\pi(B)$.*

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

$$A_i = \{X : r_{\max}(X) \leq p_i + \varepsilon \text{ and } r_{\min}(X) \geq p_i - \varepsilon\}, \quad i = 1, 2\,,$$

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

$$B = \{X : p_i + \varepsilon < r_{\max}(X) \leq p_1 + 2\varepsilon \text{ or } p_1 - \varepsilon > r_{\min}(X) \geq p_i - 2\varepsilon\}$$

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 $O\left(\frac{1}{n}\right)$. A quick calculation (relegated to the appendix) using Lemma 11 shows that

$$\pi(B) \leq \frac{e^{-\Omega(n)}}{1 - 2e^{-\Omega(n)}}\pi(A)\,.$$

This, together with Claim 13, completes the proof.

3 Conclusion

Our burn in proof in the high temperature regime shows that with high probability all the $r_G(X, e)$ are close p^*, the fixed point of $\varphi(p^*) = p^*$. A consequence is that for any collection of edges $e_1, \ldots, e_j$ the events x_{e_i} are asymptotically independent and distributed as Bernoulli(p^*). As such, the exponential random graph model is extremely 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 proves that the model fails in it's stated goal.

References

[1] Aldous, David and Fill, Jim Reversible Markov Chains and Random Walks on Graphs *Book in preparation*

[2] Park, J. and Newman, MEJ Solution of the 2-star model of a network *Physical Review E*, 70 Volume 6, 2004

[3] Park, J. and Newman, MEJ Solution for the properties of a clustered network *Physical Review E*, 2 Volume 72, 2005

[4] Chatterjee, Sourav Exact large deviations for triangles in a Random Graph Work in progress, 2008

[5] Anderson, C.J. and Wasserman, S. and Crouch, B. A p* primer: logit models for social networks *Social Networks*, 1 Volume 21: 37-66, 1999

[6] Snijders, T.A.B. and Pattison, P.E. and Robins, G.L. and Handcock, M.S. New Specifications for Exponential Random Graph Models *Sociological Methodology*, 36, Volume 1:99-153, 2006.

[7] Snijders, T.A.B. Markov chain Monte Carlo estimation of exponential random graph models *Journal of Social Structure*, 2 Volume 3: 1-40, 2002

[8] Wasserman, S. and Pattison, P. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p *Psychometrika*, 3 Volume 61: 401-425, 1996

[9] Frank, O. and Strauss, D Markov Graphs *Journal of the American Statistical Association* 395, Vol 81: 832-842, 1986

[10] Newman, ME The structure and function of complex networks *SIAM Review*, 2, Volume 45: 167–256, 2003

[11] Albert, R. and Barabási, A.L. Statistical mechanics of complex networks *Reviews of Modern Physics*, 1 Volume 74: 47-97, 2002

[12] Durrett, R. Random Graph Dynamics *Cambridge University Press*, 2006

[13] Mendes, JFF and Dorogovtsev, SN Evolution of networks *Oxford University Press, Oxford*, 2003

[14] M. Dyer, A. Frieze, M. Jerrum. On Counting Independent Sets In Sparse Graphs *SIAM J. of Computing*, Vol 31, No. 5, pp. 1527-1541.

[15] Bubley, R. and Dyer, M. Path coupling: A technique for proving rapid mixing in Markov chains FOCS 1997, 223-231.

[16] Dyer, M. and Frieze, A. Randomly coloring graphs with lower bounds on girth and maximum degree, *Random Structures and Algorithms*, Vol 23, No 2: 167-179, 2003

Appendix

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 14 *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

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

Proof of Claim 9

For ease of notation, we suppress the dependence of $r_{\max}$ on the configuration X. 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'.

$$\begin{aligned}&\mathbb{E}\left[\frac{N_G(X(1),e)-N_G(X(0),e)}{n^{|V|-2}}\right]\\ &= \frac{1}{n^{|V|-2}}\left[-\binom{n}{2}^{-1}(|E|-1)N_G(X,e) + \binom{n}{2}^{-1}\sum_{e'\neq e} N_G(X,e,e')\mathbf{P}(X_{e'}(1)=1|e' \text{ updated})\right].\end{aligned} \tag{10}$$

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

$$\begin{aligned}\mathbf{P}(X_{e'}(1)=1|e' \text{ updated}) &= \frac{\exp(\partial_e H(X))}{\exp(\partial_e H(X))+1}\\ &= \frac{\exp\left(\sum_i \beta_i \frac{N_{G_i}(X,e)}{n^{|V|-2}}\right)}{\exp\left(\sum_i \beta_i \frac{N_{G_i}(X,e)}{n^{|V|-2}}\right)+1}\\ &\leq (1+o(1))\frac{\exp\left(\sum_i \beta_i \frac{N_G(K_n,e)(r_{\max})^{|E_i|-1}}{n^{|V|-2}}\right)}{\exp\left(\sum_i \beta_i \frac{N_G(K_n,e)(r_{\max})^{|E_i|-1}}{n^{|V|-2}}\right)+1}\\ &= (1+o(1))\varphi(r_{\max})\,.\end{aligned} \tag{11}$$

Next, by Lemmas 14 and 7 and the definition of $r_{\max}$, we have

$$\begin{aligned}\sum_{e'} N_G(X,e,e') &= \sum_{\alpha} N_{G_\alpha}(X,e) \\ &\le \sum_{\alpha} N_{G_\alpha}(K_n,e)(r_{\max})^{|E|-2} \\ &= \sum_{e'\neq e} N_G(K_n,e,e')(r_{\max})^{|E|-2} \\ &= |E|(|E|-1)2n^{|V|-2}(r_{\max})^{|E|-2}(1+o(1))\,.\end{aligned} \tag{12}$$

Using the estimates (11) and (12), equation (10) gives

$$\begin{aligned}&\mathbb{E}\left[\frac{N_G(X(1),e)-N_G(X(0),e)}{n^{|V|-2}}\right] \\ &\le (1+o(1))\frac{1}{n^{|V|-2}\binom{n}{2}}\left[-(|E|-1)N_G(X,e)+\varphi(r_{\max})2|E|(|E|-1)n^{|V|-2}(r_{\max})^{|E|-2}\right] \\ &= (1+o(1))\frac{1}{n^{|V|-2}\binom{n}{2}}\left[-(|E|-1)2|E|n^{|V|-2}r(G,e)^{|E|-1}+\varphi(r_{\max})2|E|(|E|-1)n^{|V|-2}(r_{\max})^{|E|-2}\right] \\ &= (1+o(1))\frac{2}{\binom{n}{2}}|E|(|E|-1)[-r(G,e)^{|E|-1}+\varphi(r_{\max})(r_{\max})^{|E|-2}]\,.\end{aligned}$$

Proof of Lemma 8

The lemma is proved by coupling each of the random variables $N_G(X(t),e)$ (one for each edge e and graph G), with an independent biased random walk.

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

$$(r-2\delta)^{|E|-1} > \varphi(r+\delta)(r+\delta)^{|E|-2}+\varepsilon\,. \tag{13}$$

It follows by Claim 9 that if $r_G(X(t),e)\ge r_{\max}(X(0))-2\delta$ and $r_{\max}(X(t))\le r_{\max}(X(0))+\delta$, then for sufficiently large n,

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

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

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

and put

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

and

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

$B_{t_1,t_2}(e,G,\delta)$ 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, $1\le\tau\le T$, is contained in the event $\bigcup_{e,G}\bigcup_{0\le t_1<t_2\le T} B_{t_1,t_2}(e,G,\delta)$. 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 $\bigcup_{0\le t_1<t_2\le T} B_{t_1,t_2}(e,G,\delta)$ is bounded as*

$$\mathbf{P}\left(\bigcup_{0\le t_1<t_2\le T} B_{t_1,t_2}(e,G,\delta)\right)\le e^{-\Omega(n)}\,. \tag{14}$$

Proof.

For all X we have $N_{G_i}(X,e,e')\le N_{G_i}(K_n,e,e')$. 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

$$N_{G_i}(X,e,e')=O(n^{-1}). \tag{15}$$

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

$$N_{G_i}(X,e,e')=O(n^{-2}). \tag{16}$$

Although the claim concerns the random variable $r(G,e)$, 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

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

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

$$B_{t_1,t_2}(e,G,\delta)\subseteq\{S_{t_1,t_2}\ge\delta/2\}\,. \tag{17}$$

We have

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

From Claim 9 and equation (13) 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 $|Y_t-Y_{t-1}|=O(n^{-1})$ with probability $O(n^{-1})$ and otherwise $|Y_t-Y_{t-1}|=O(n^{-2})$.

Thus we have

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

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

$$\mathbb{E}\left(e^{\theta(Y_{t_2}-Y_{t_2-1}+\frac{\gamma}{2n^2})\mathbb{1}(D_{t-1}(e,G,\delta))}|\mathcal{F}_{t-1}\right)\leq 1$$

and so

$$
\begin{aligned}
\mathbb{E}\left[e^{\theta S_{t_1,t_2}}\right]&\leq\mathbb{E}\left[e^{\theta S_{t_1,t_2-1}}\right]\\
&\leq 1\,,
\end{aligned}
$$

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 $\{G_\lambda\}$,

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

This gives the estimate

$$\mathbf{P}(S_{t_1,t_2}\geq\alpha)\leq e^{-c\alpha n}\mathbb{E}\left[e^{\theta(Y_t-Y_0)}\right]=e^{-\Omega(n)}\,. \tag{18}$$

and so

$$P(r_G(X(t_2),e)-r_G(X(t_1),e)>\delta/2)=e^{-\Omega(n)}. \tag{19}$$

We may now apply (19) to equation (17), resulting in

$$\mathbf{P}\left(\bigcup_{0\leq t_1\leq t\leq t_2\leq T}B_{t_1,t_2}(e,G,\delta)\right)\leq T^2e^{-\Omega(n)}(1+o(n))=e^{-\Omega(n)}\,, \tag{20}$$

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, $r_G(X(0),e)\geq r_{\max}-2\delta$. Then for $T=cn^2$,

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

where the last step follows since each of the T increments in $S_{1,T}$ contribute $\gamma/2n^2$ on the event $\cap_{1\le t\le T} D_t(e,G,\delta)$. Choosing $c \ge 3/\gamma$ and using the estimate on the deviation of S_{t_1,t_2} (19) gives

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

Finally, we have

$$\begin{aligned}
&\mathbf{P}(r_G(X(T),e) \ge r_{\max} - \delta) \\
&\le \mathbf{P}(r_G(X(T),e) \ge r_{\max} - \delta, r_G(X(t),e) < r_{\max} - 2\delta \text{ for some } t \in [1,T]) + e^{-\Omega(n)} \\
&\le \mathbf{P}(\cup_{1\le t_1\le T} B_{t_1,T}(e,G,\delta)) + e^{-\Omega(n)} \le e^{-\Omega(n)}.
\end{aligned}$$

The union bound on probabilities applied over the set of edges e and graphs G completes the proof of Lemma 8.

Proof of Lemma 12

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

$$P(X^{\pm}_{e'}(1) = 1) = \frac{\exp(\partial_{e'} H(X^{\pm}(0)))}{1 + \exp(\partial_{e'} H(X^{\pm}(0)))}. \tag{21}$$

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 14 and equation (8) we have that for large enough n,

$$\partial_{e'} H(X^{\pm}(0)) \le \sum_{i=1}^{s} \frac{\beta_i (p^* + \epsilon)^{|V_i|-2} N_{G_i}(K_n, e')}{n^{|V_i|-2}} = \Psi(p^* + \epsilon)(1 + o(1)). \tag{22}$$

and similarly

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

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

$$\left.\frac{d}{dx}\frac{e^x}{1+e^x}\right|_{\partial_{e'} H(X^+(0))} \le (1+\epsilon') \left.\frac{d}{dx}\frac{e^x}{1+e^x}\right|_{\Psi(p^*)}. \tag{23}$$

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

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

where the second and fourth lines follow from Lemma 7 and the inequality follows from equation (8). By Lemma 14 we have that

$$\sum_{e' \neq e} \partial_e \partial_{e'} H(X^+(0)) \leq (1+o(1)) \sum_{i=1}^{s} 2|E_i|(|E_i|-1)(p^*+\epsilon)^{|E_i|-2} = (1+o(1))\Psi'(p^*+\epsilon) \tag{24}$$

By Taylor series for small h we have that $\frac{e^{x+h}}{1+e^{x+h}} - \frac{e^x}{1+e^x} \leq \frac{d}{dx}\frac{e^x}{1+e^x}\Big|_x (h + O(h^2))$ and so using equation (21),

$$P(X^+_{e'}(1)=1) - P(X^-_{e'}(1)=1) = \frac{d}{dx}\frac{e^x}{1+e^x}\Big|_{\partial_{e'}H(X^+(0))} \cdot \left(\partial_e \partial_{e'} H(X^+(0)) + O((\partial_e \partial_{e'} H(X^+(0))^2)\right) \tag{25}$$

$$\leq (1+\epsilon')(1+o(1))\partial_e \partial_{e'} H(X^+(0)) \; \frac{d}{dx}\frac{e^x}{1+e^x}\Big|_{\Psi(p)} \tag{26}$$

using equations (23) and the fact that by equation (15) 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 (25) and (24) that for any $\epsilon'' > 0$

$$\begin{aligned} Ed_H(X^+(1), X^-(1)) &\leq 1 - \binom{n}{2}^{-1} \left[1 - \sum_{e' \neq e} (1+\epsilon')(1+o(1))\partial_e \partial_{e'} H(X^+(0)) \; \frac{d}{dx}\frac{e^x}{1+e^x}\Big|_{\Psi(p)}\right] \\ &\leq 1 - \binom{n}{2}^{-1} \left[1 - (1+\epsilon')(1+o(1))\Psi'(p^*+\epsilon) \; \frac{d}{dx}\frac{e^x}{1+e^x}\Big|_{\Psi(p)}\right] \\ &\leq 1 - \binom{n}{2}^{-1} \left[1 - (1+\epsilon'')(1+o(1))\varphi'(p^*)\right] \end{aligned}$$

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

Proof of slow mixing for local dynamics

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

$$A_i = \{X : r_{\max}(X) \leq p_i + \varepsilon \text{ and } r_{\min}(X) \geq p_i - \varepsilon\}, \quad i = 1, 2\,,$$

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

$$B = \{X : p_i + \varepsilon < r_{\max}(X) \leq p_1 + 2\varepsilon \text{ or } p_1 - \varepsilon > r_{\min}(X) \geq p_i - 2\varepsilon\}$$

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 $O\left(\frac{1}{n}\right)$.

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 8 holds. Then

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

and

$$\mathbf{P}(X(t) \in B | X(0) \in A \cup B) = e^{-\Omega(n)}\,. \tag{28}$$

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 (27) we have

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

Similarly, using (28),

$$\begin{aligned}\mathbf{P}(X(t) \in B, X(0) \in A \cup B) &= \mathbf{P}(X(t) \in B | X(0) \in A \cup B) \mathbf{P}(X(0) \in A \cup B) \\ &\leq e^{-\Omega(n)} \mathbf{P}(X(0) \in A \cup B) \\ &= e^{-\Omega(n)} (\mathbf{P}(X(0) \in A) + \mathbf{P}(X(0) \in B))\,.\end{aligned} \tag{30}$$

Combining (29) and (30), we have

$$\begin{aligned}\pi(B) &= \mathbf{P}(X(t) \in B) \\ &= \mathbf{P}(X(t) \in B, X(0) \in C) + \mathbf{P}(X(t) \in B, X(0) \in A \cup B) \\ &\leq e^{-\Omega(n)} (\mathbf{P}(X(0) \in A) + 2\mathbf{P}(X(0) \in B)) \\ &= e^{-\Omega(n)} (\pi(A) + 2\pi(B))\,,\end{aligned} \tag{31}$$

which, upon rearranging, gives

$$\pi(B) \leq \frac{e^{-\Omega(n)}}{1 - 2e^{-\Omega(n)}} \pi(A)\,. \tag{32}$$

Together with Claim 13 this completes the proof.