A Review of the Mixing Time of Exponential Random Graphs

Soumya Basu,
University of Texas at Austin, USA

***Abstract*—In recent years various random graph models have been introduced to study real world networks, e.g. social, telecommunication, vehicular networks. The exponential random graph model promises to capture the reciprocity exhibited by these networks. In dealing with parameter estimation, hypothesis testing and similar practical issues the sampling from this distribution is necessary. The sampling is typically based on Markov chain Monte Carlo methods. Specifically, Glauber dynamics or Metropolis-Hastings method.**

In this paper we review the seminal work on Mixing Time of Random Graphs by Shankar Bhamidi et al. The mentioned work differentiates the exponential random graph model in high and low temperature regimes. In high temperature employing Glauber dynamics the mixing in $\Theta(n^2 logn)$ is proved, n being the number of vertices in the graph. But on the contrary, in the low temperature regime the mixing is exponentially slow with the use of $O(n)$ Glauber dynamics in each step. Furthermore, they showed that in the high temperature regime the model becomes similar to Erdős-Rényi Graphs and fails to model the desired reciprocity.

I. Introduction

In the recent past the study of real world networks, specifically social, biochemical, vehicular and web based networks, have gained popularity leading to the development of various mathematical models. These models have focused on explaining the typical characteristics, such as power law degree behaviour, small world properties and clustering. An important example is the clustering in social networks which emerges from the reciprocity of friendship among people. Many otherwise useful network models, e.g. preferential attachment, fail to model clustering given their locally tree like structures.

The exponential random graph model follows the statistical mechanics approach which is to weight a probability measure on the space of graphs by defining a Hamiltonian. It uses parametric modelling to assign higher probability mass to the entities with some "desirable" properties. The Hamiltonian in this model is a function of subgraph counts of special structures which will be formally introduced in Section III. The model can be efficiently used in parameter estimation, hypothesis testing or modelling, given specific observations over subgraph counts. For this purpose efficient sampling of the graphs is necessary. Markov chain Monte Carlo (MCMC) techniques, specifically Glauber dynamics or Metropolis-Hastings method, forms reversible ergodic Markov chains leading to the convergence to the required stationary distribution.

This paper mainly reviews the contribution of S. Bhamidi, G. Brestler, A. Sly on the mixing time of exponential random graphs. We present the main results and give adequate outline to the proofs presented in the paper [1]. Additionally, we try to develop some valuable insights into the overall approach taken in their work. The rest of the paper is organised as follows. The related works are presented briefly in Section II. In Section III we formally define the model and associated MCMC sampling. The phases of the graph is explained in this section. Next, we present the main results by S. Bhamidi et al. in Section IV. Section V introduces the basic insights behind the proof and gives a sketch of the proof. In Section VI the proofs are presented with the help of various lemmas. We give an outline of the proofs,whereas the detailed proofs can be found in [1]. Finally, we conclude with a few possible directions in Section VII.

II. Related Work

Frank and Strauss in [2] proposed Markov graphs as a family of random graphs parametrized over the subgraph counts of stars and triangles. More specifically, given a fixed number of nodes, non-incident edges are independent conditional on the rest of the graph. Wasserman and Pattinson [3] extended the modelling to subgraph counts of general graphs. Based on this model several empirical level results has been produced applying Markov chain algorithms based on parameter values. Discussions can be found in [4]. The study on mixing time of graphs with only stars and triangles can be found in the works by Mark Newman *et al.* [5]. The paper by Shankar Bhamidi *et al.* [1] presents an extensive study in the mixing time of the generic model.

III. Exponential Random Graphs: Model and Characterization

In this section we formally present the model for exponential random graph and the MCMC techniques involved. The different phases of the model is defined in this section which will help characterize the mixing in these graphs.

A. Notation

$\mathcal{G}_n$ is the space of all graphs with n vertices and vertex set $[n] := 1, 2, ..., n$. Any graph $X \in \mathcal{G}_n$ is given by the vector (x_e), $\forall\ e = (i, j)$ where $x_e = \mathbb{1}(e \in X)$. Denote, $X_{e+} = X \cup \{e\}$ and $X_{e-} = X \setminus \{e\}$. Let $S_X(m)$ be the set of all possible subgraphs of X with m vertices and $Y \cong G$ mean, Y contains G.

Given a graph G with m vertices define the following subgraph counts.
The subgraph count of G in X

$$N_G(X) = |\{Y : Y \in S_X(m) \wedge Y \cong G\}|\,.$$

The subgraph count of G in $X \cup \{e\}$ containing $\{e\}$

$$N_G(X,e) = \left|\{Y : Y \in S_{X\cup\{e\}}(m) \wedge Y \cong G\}\right|.$$

The subgraph count of G in $X \cup \{e, e^{'}\}$ containing $\{e, e^{'}\}$

$$N_G(X,e,e^{'}) = \left|\{Y : Y \in S_{X\cup\{e,e'\}}(m) \wedge Y \cong G\}\right|.$$

B. Gibbs Measure

We define the probability measure on the space $\mathcal{G}_n$ as follows.
Fix $s \geq 1$ and fix "desired" structures $G_1, G_2, ..., G_s$ with $G_i(V_i, E_i)$ as undirected graphs and $\max|V_i| \leq L$. By convention G_1 is a single edge. The vector $\boldsymbol{\beta} = \{\beta_i\}_1^s \in \mathcal{B}$ is the parameter for the model, where $\mathcal{B} = \mathbb{R} \times \mathbb{R}_{++}^{s-1}$.

Definition For $G_1, G_2, ..., G_s$ and $\boldsymbol{\beta}$ as above, the Gibbs measure in the space $\mathcal{G}_n$ is defined as the probability measure

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

Here $Z_n(\beta)$ is the partition function and note that the scaling by $n^{|V_i|-2}$ makes the contribution of G_i of the order n^2 $\forall i$.

The term in the exponent is refered as the *Hamiltonian*

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

Note that $H(X) : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}_+$ is a function of all the edge indicators, $X(e)$. Using Fourier decomposition over the basis functions $\Pi_{e\in\mathcal{S}} X(e)$, where $\mathcal{S}$ is any subset of edges, we have the decomposition

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

Here, $A_e(\cdot)$ consists of all terms dependent of edge e and $B_e(\cdot)$ denotes all independent terms. Furthermore, we have $\partial_e H(X) = A_e(X_{e+})$.

C. Glauber Dynamics and Local Chains

Glauber dynamics is one of the many ways of defining a Markov chain given a state space.

Definition The Glauber dynamics is a ergodic reversible Markov chain defined on $\mathcal{G}_n$ with stationary probability $p_n(\cdot)$ and the following transitions,
Given current state X,

1) choose an edge e, uniformly at random,
2) move to next state $X^{'} = X_{e\pm}$
   w.p. proportional to $p_n(X_{e\pm})$.

Call a chain on $\mathcal{G}_n$ *Local chain* if $O(n)$ edges (following Glauber dynamics) are updated in one step.

We also give the definition of mixing time here for the sake of completeness.

Definition For Markov chain the mixing time $t_{mix}(\epsilon)$ is the number of steps required to guarantee that the chain, starting from any arbitrary state, is within total variation distance ϵ from the stationary distribution.

The value of ϵ is generally taken to be e^{-1} or $1/4$. We consider $\epsilon = o(1)$ to show mixing.

Lemma III.1 ([1]). *Given that we choose 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 $\frac{1}{1+exp(\partial_e H(X))}$.*

D. Phases in Exponential Random Graphs

With the above definitions, for a fixed $\boldsymbol{\beta}$ define,

$$\Psi_{\boldsymbol{\beta}}(p) = \sum_{i=1}^s 2\beta_i |E_i| p^{|E_i|-1}$$
$$\varphi_{\boldsymbol{\beta}}(p) = \frac{\exp\left(\Psi_\beta(p)\right)}{1+\exp\left(\Psi_\beta(p)\right)}$$

Note that $\Psi_{\boldsymbol{\beta}}$ is smooth over $p \in [0,1]$. Further, $\varphi_{\boldsymbol{\beta}}(0) > 0$ and $\varphi_{\boldsymbol{\beta}}(1) < 1$ implies $\exists p^* \in (0,1)$ such that $\varphi_{\boldsymbol{\beta}}(p^*) = p^*$. We denote p^* as a fixed point of $\varphi_{\boldsymbol{\beta}}(p)$. We use $\varphi_{\boldsymbol{\beta}}(p)$ and $\varphi(p)$ interchangeably in the rest of the paper.

1) High temperature phase: An exponential random graph is in the high temperature phase if $\beta \in \mathcal{B}_H$. Where $\mathcal{B}_H \subseteq \mathcal{B}$,

$$\mathcal{B}_H = \{\boldsymbol{\beta} | \exists!\ p^* \text{ s.t. } 0 < \varphi_{\boldsymbol{\beta}}^{'}(p^*) < 1\}.$$

2) Low temperature phase: An exponential random graph is in the low temperature phase if $\beta \in \mathcal{B}_L$. Where $\mathcal{B}_L \subseteq \mathcal{B}$,

$$\mathcal{B}_L = \{\boldsymbol{\beta} | \exists \text{ min two } p^* \text{ s.t. } 0 < \varphi_{\boldsymbol{\beta}}^{'}(p^*) < 1\}.$$

3) Critical phase: An exponential random graph is in the critical phase if $\beta \notin \mathcal{B}_H \cup \mathcal{B}_L$. This happens if one of the fixed point is also an inflection point of $\varphi(p)$. That are specifically $(s-1)$-dimensional manifold in the intersection of closure of $\mathcal{B}_H$ and closure of $\mathcal{B}_L$.

IV. Results: Mixing Time Bounds and Asymptotic Independence of Edges

The first two results show that high and low temperature phases determine the mixing time with local chains.

Theorem IV.1 (**High Temperature** [1]). *If $\varphi(p)$ is in the high temperature regime then the mixing time of the Glauber dynamics is $\Theta\left(n^2 \log n\right)$.*

Theorem IV.2 (**Low Temperature** [1]). *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 Glauber dynamics but for any local dynamics on $\mathcal{G}_n$.*

Theorem IV.3 (**Asymptotic Independence of Edges** [1]). *Let X be drawn from the exponential random graph distribution in the high temperature phase. Let $e_1, ..., e_k$ be an arbitrary collection of edges with the associated indicator random variables $x_{e_i} = \mathbb{1}(e_i \in X)$. Then for any $\epsilon > 0$, there is an n such that for all$(a_1, ..., a_2) \in \{0,1\}^k$ the random variables $x_{e_1}, ..., x_{e_k}$ satisfy*

$$\left|\mathbf{P}(x_1 = a_1, ..., x_k = a_k) - (p^*)^{\sum a_i}(1-p^*)^{k-\sum a_i}\right| \leq \frac{\epsilon}{n^{|V|}}.$$

This theorem shows in high temperature phase though sampling is possible, the exponential random graph behaves similar to Erdős-Rényi graph, $G(n, p^*)$.

V. Proof Intuition and Outline

In this section we present the importance of the phases in mixing time, through the motivation derived from Erdős-Rényi model. Next with this intuition we give a brief outline for the proof of main results.

A. Insights through Erdős-Rényi model

The function defining the phases in the exponential random graph have the motivation that if we choose X from $G(n,p)$ then all the edge update probabilities $\frac{exp(\partial_e H(X))}{1+exp(\partial_e H(X))}$ are approximately $\varphi(p)$.

Assume our initial choice is from some $G(n, p_{init})$. Note if $p_{init} = p^*$ the edge update process given by Glauber dynamics is same as the edge update process in $G(n,p^*)$ with high probability. Additionally, if we have $\varphi^{'}(p^*) < 1$, observe that starting from $p_{init} = p^* + \epsilon$ results in $\varphi(p^* + \epsilon) < p^* + \epsilon$ for small $\epsilon > 0$. This means after a single step of Glauber dynamics $X^{'}$ becomes more like $G(n,p^*)$. Similar result holds true for $p_{init} = p^* - \epsilon$, as well. This gives a drift towards the fixed points with $\varphi^{'}(p^*) < 1$. On the contrary, if we have $\varphi^{'}(p^*) > 1$ then we can observe a drift away from the fixed point p^*. Call these type of fixed points, *unstable* points and the previous ones, *stable* points.

Furthermore, if β is not in the critical phase then the stable and unstable points occur alternatively. Therefore, (intuitively) if we start in between two *unstable* points, we will drift towards the *stable* point in the interior of these two points. Therefore, the *unstable* points divide the interval $[0,1]$ into *modes*. Specifically, in high temperature there exists one such *mode* which gives some implication of fast mixing as well as the similarity to $G(n,p^*)$. Whereas in low temperature the existence of multiple such *modes* lead to slow mixing.

Next we give a notion of closeness of a random graph X to certain $G(n,p)$. Define

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

This is (asymptotically) the maximum likelihood choice for parameter p of $G(n,p)$, having observed $N_G(X,e)$ subgraphs of G containing the edge e. Further define for each fixed point

$$\mathbf{G} = \{X : r_G(X,e) \in (p^* - \epsilon, p^* + \epsilon);\ \forall G \in G_\lambda, e \in E\}$$

where $\{G_\lambda\}$ is the set of all graphs with at most L vertices. Observe, if $X \in \mathbf{G}$, for ϵ small enough $X \sim G(n,p^*)$ approximately.

B. Proof outline

Here we present a sketch of the proofs of main results with the motivation presented in the previous section.

- We first show that for any $G \in \{G_\lambda\}$ and edge e, the Glauber dynamics produce drift in the variables $r_G(X,e)$ from $[p_0 + \mu, p_1 - \epsilon]$ and $[p_1 + \epsilon, p_2 - \mu]$ towards $[p_1 - \epsilon, p_1 + \epsilon]$ where p_0, p_2 are *unstable* points and p_1 is the *stable* point in the interior for any $\epsilon, \mu > 0$. We call the initial phase Burn-in period, when the $r_G(X,e)$ are not contained in $\mathbf{G}$ with high probability (w.h.p.).
- Next we prove that in Burn-in period we achieve $o(1)$ drift in $O(n^2)$ steps w.h.p. Further, once some $r_G(X,e)$ is inside $\mathbf{G}$ it can be shown that the flow outside this region is exponentially slow, i.e. $e^{-\Omega(n)}$.
- Taking into consideration the monotonicity of the Gibbs measure we have standard monotone coupling in the state space. With the technique of path coupling [6] we show that fast mixing occurs in high temperature phase.
- In low temperature phase, we have multiple *modes* each with their own region $\mathbf{G}$. Now using the fact the flow out of $\mathbf{G}$ is exponentially small we use a known conductance bound to show exponential mixing in this phase.
- Finally using the containment of the variables $r_G(X,e)$ inside $\mathbf{G}$, we can prove the asymptotic independence of edges in high temperature phase.

VI. Proof of Main Results

In this section we first introduce the necessary lemmas to prove the main results. Then we give the proofs of main results as an outline highlighting the key points. For detailed proofs refer to [1].

A. Subgraph Counts

We present a few simple lemmas on subgraph count.

Lemma VI.1 ([1]). *Consider the complete graph on n vertices, K_n. We have,*

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

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

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

Lemma VI.2 ([1]). *For an edge α in the graph G, let $G_\alpha = G \setminus \{\alpha\}$. Then*

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

B. Burn-in period

The key to the proofs is the containment of $r_G(X,e)$ in $\mathbf{G}$ after a short time interval, burn-in period. From hereon, $X \equiv X(0)$ and $r_G(X,e) \equiv r(G,e)$. Define, $r_{max} = \max_{e, G \in G_\lambda} r_G(X,e)$ and $r_{min} = \min_{e, G \in G_\lambda} r_G(X,e)$.

Lemma VI.3 ([1]). *The expected change in $N_G(X,e)$ after one step of the Glauber dynamics, starting from any configuration X, can be bounded as given in the equation (*).*

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

Proof. After one step, starting from $X(0)$
$\mathbb{E}(N_G(X(1),e) - N_G(X(0),e))$
$= -\binom{n}{2}^{-1}(|E|-1)N_G(X,e)$
$+\binom{n}{2}^{-1}\sum_{e'\neq e} N_G(X,e,e')\mathbb{P}(X_{e'}(1)=1|e' \textit{updated})$.

The first part is the expected loss due to removing the random edge e' and the second part is the expected gain from adding the random edge e'. Using Lemma III.1, definition of r_{max} and $N_G(X,e) = 2|E|n^{|V|-2}r(G,e)^{|E|-1}$ the result follows. □

Lemma VI.4 ([1]). *Let $\varphi(p^*) = p^*$, $\varphi'(p^*) < 1$ and $\bar{p}$ be the least solution greater than p^*, if such a solution exists, or 1 otherwise. Let starting from $X(0)$ for some $\mu > 0$, $p^* + \mu \leq r_{max}(X(0)) \leq \bar{p} - \mu$. Then $\exists\ \delta, c > 0$ depending only on μ, L and φ so that after $T = cn^2$ steps of the Glauber dynamics, it holds that $r_{max}(X(T)) \leq r_{max}(X(0)) - \delta$ with probability $1 - e^{-\Omega(n)}$.*

Proof. The key idea is to couple each random variable $N_G(X,e)$ with a biased random walk. Note for $\varphi'(p^*) < 1$ $\exists \epsilon, \delta > 0$ such that for any $r \in [p^* + \mu, \bar{p} - \mu - \delta]$,

$$(r-2\delta)^{|E|-1} > \varphi(r+\delta)(r+\delta)^{|E|-2} + \epsilon. \qquad (**)$$

Define events, $R_t(G,e,\delta) = \{r_G(X(t),e) \geq r_{max} - 2\delta\}$, $A_t(\delta) = \{r_{max}(X(t)) \leq r_{max} + \delta\}$ and the "appropriate" edge statistics $D_t(G,e,\delta) = R_t(G,e,\delta) \cap A_t(\delta)$.

From Lemma VI.3 and equation (**), if $r_G(X(t),e) \in D_t(G,e,\delta)$ then, for large n we have negative expected drift of $N_G(X(t),e)$

$$\mathbb{E}\left[\frac{N_G(X(t+1),e) - N_G(X(t),e)}{n^{|V|-2}}\mathbb{1}(D_t(e,G,\delta))\right] \leq -\frac{\gamma}{n^2}.$$

Where, γ depends only on φ,δ and ϵ.

We claim the following about the event,

$$B_{t_1,t_2}(e,G,\delta) = (\cap_{t_1\leq t<t_2} \cup_{e,G} D_t(e,G,\delta)) \cup \{r_G(X(t_2),e) - r_G(X(t_1),e) > \delta/2\}.$$

Claim VI.5 ([1]). *The probability of the event $\cup_{0\leq t_1<t_2\leq T} B_{t_1,t_2}(e,G,\delta)$ is bounded*

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

Proof. Define a new random variable, S_{t_1,t_2}, which is the sum of the one step drift of $N_G(X(t),e)$ for $t_1 \leq t < t_2$, with an added bias of $\frac{\gamma}{2n^2}$. Precisely,

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

For $\theta = cn$ and sufficiently small $c > 0$, we have

$$\mathbb{E}\left(e^{\theta S_{t_1,t_2}}|\mathcal{F}_{t_2-1}\right) \leq e^{\theta S_{t_1,t_2-1}}.$$

For $\alpha < sup_{x\in[p^*,1],G\in\{G_\lambda\}}\{(x+\delta/2)^{|E|-1} - x^{|E|-1}\}$ applying Chernoff bound gives, $\mathbf{P}(S_{t_1,t_2} \geq \alpha) \leq e^{-\Omega(n)}$. Now notice that $B_{t_1,t_2}(e,G,\delta) \subseteq \{S_{t_1,t_2} \geq \delta/2\}$. Finally by applying union bound over t_1 and t_2 for $T = poly(n)$

$$\mathbf{P}\left(\cup_{0\leq t_1<t_2\leq T} B_{t_1,t_2}(e,G,\delta)\right) \leq T^2 e^{-\Omega(n)}(1+o(n)).$$

□

The probability of $\{r_G(X(t),e) \geq r_{max} - 2\delta\}$ for $1 \leq t \leq T$ is bounded as

$$\mathbf{P}\left(r_G(X(t),e) \geq r_{max} - 2\delta, 1 \leq t \leq T\right) \leq \mathbf{P}\left(S_{1,T} \geq -1 + \frac{T\gamma}{2n^2}\right) + e^{-\Omega(n)} \leq e^{-\Omega(n)}.$$

In $T \geq \frac{3n^2}{\gamma}$ steps all the random variables $r_G(X(T),e)$ reach below $r_{max} - 2\delta$ w.h.p.

Claim VI.5 also shows that with probability $e^{-\Omega(n)}$ the event $\cup_{1\leq\tau\leq T}\{r_G(X(\tau),e) \geq r_{max} + \delta\}$ occurs. Therefore, applying Claim VI.5, conditioned on $r_G(X(t),e) < r_{max} - 2\delta$ for some $1 \leq t \leq T$, we have the bound

$$\mathbf{P}(r_G(X(T),e) \geq r_{max} - \delta) \leq e^{-\Omega(n)}.$$

Applying union bound on G and e proves the lemma.

□

Corollary VI.6. *In high temp $\exists c, \epsilon > 0$ such that for all G, e, for any $X(0) = x$ and $t \geq cn^2$,*

$$\mathbf{P}\left(|r(X(t)) - p^*| \geq \epsilon\ |X(0) = x\right) \leq e^{-\Omega(n)}$$

Corollary VI.7. *In low temp for $\varphi'(p^*) < 1$, $\exists\ \epsilon, \alpha > 0$ such that for all G, e and for $X(0)$ with $|r(X(0) - p^*| \leq \epsilon$,*

$$\mathbf{P}\left(\sup_{0<t<e^{\alpha n}} |r(X(t)) - p^*| \geq 2\epsilon\ |X(0) = x\right) \leq e^{-\Omega(n)}$$

C. Proof of lower bound on mixing time (Theorem IV.2) [1]

We apply a known conductance bound to give a lower bound on the mixing time in low temperature phase.

Lemma VI.8 ([7]). *Let $\mathcal{M}$ be a Markov chain with state space Ω, transition probability matrix P and stationary distribution π. Let $A \subset \Omega$ with $\pi(A) \leq \frac{1}{2}$ and $B \subset \Omega$ be the "barrier" such that $P_{ij} = 0\ \forall i \in A \setminus B,\ j \in A^c \setminus B$.Then the mixing time $t_{mix} \geq \pi/8\pi(B)$.*

We define the space $\Omega = \mathcal{G}_n$. The transitions, p_{ij}, are defined according to the Glauber dynamics and the stationary distribution π is given by the Gibbs measure.

In low temperature, $\exists p_1, p_2$ and $\epsilon > 0$ such that $\varphi(p_i) = p_i$ and $\varphi'(p_i) < 1$ and $\varphi(p) > p$ for $p \in [p_i - 3\epsilon, p_i)$, $\varphi(p) < p$ for $p \in (p_i, p_i + 3\epsilon)$ for $i \in \{1,2\}$. Further, define the sets $A_i = \{X : |r(X) - p_i| \leq \epsilon\}$ for $i \in \{1,2\}$ and without loss of generality assume $\pi(A_1) \leq \frac{1}{2}$.

Observe that r_{max} and r_{min} can have $O(1/n)$ change by one edge update. Therefore, for local chains, i.e. $o(n)$ edge updates, the set $B = \{X : p_1 + \epsilon < r_{max}(X) \leq p_1 + 2\epsilon \vee p_1 - \epsilon > r_{min}(X) \geq p_1 - 2\epsilon\}$ forms a "barrier" for A_1. Denote, $A = A_1$.

For $C = A^c \setminus B$ and $t = cn^2$ we have from Lemma VI.4

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

Drawing $X(0)$ according to π we have $X(t) \sim \pi$. Therefore,

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

Through simple algebra and application of Lemma VI.8 we get $t_{mix} \geq \frac{\pi(A)}{8\pi(B)} \geq e^{\Omega(n)}$. □

D. Proof of upper bound on mixing time (Theorem IV.1) [1]

We begin the proof with a lemma showing the negative drift using monotone coupling.

Lemma VI.9 ([1]). *Let $p^* \in [0,1]$ be a solution to $\varphi(p) = p$ and $\varphi^{'}(p^*) \in (0,1)$. Suppose that $X^+(0) \geq X^-(0)$ and they differ in exactly one edge e. There exists $\epsilon, \delta > 0$ such that if $\forall G \in \{G_\lambda\}$ and $\forall e^{'} \in E(G)$, $|r(G, e^{'}) - p^*| < \epsilon$ holds, then for large enough n, a single step of Glauber dynamics can be coupled such that $\mathbb{E}d_H(X^+(1), X^-(1)) \leq 1 - \delta n^{-2}$, where d_H is the Hamming distance.*

Proof. Take the standard monotone coupling. Choose $e^{'}$, which is to be updated by the Markov chain, uniformly at random.

First we give the following bounds,

$$\partial_{e'} H(X^{\pm}(0)) \leq \frac{1}{n^{|V_i|-2}} \sum_{i=1}^{s} \beta_i (p^* + \epsilon)^{|V_i|-2} N_i(K_n, e^{'}),$$
$$(1 - o(1))\Psi(p^* - \epsilon) \leq \partial_{e'} H(X^{\pm}(0)) \leq \Psi(p^* + \epsilon),$$
$$\sum_{e^{'} \neq e} \partial_e \partial_{e'} H(X^+(0)) \leq (1 + o(1))\Psi^{'}(p^* + \epsilon). \qquad (\#)$$

The first bound uses the monotonicity of the Gibbs measure and the other bounds follow in the same line.

Further, using the Lemma III.1 and observing that $\partial_{e'} H(X^-(0) = \partial_{e'} H(X^+(0)) - \partial_e \partial_{e'} H(X^+(0))$, we have

$$\begin{aligned}&\mathbf{P}(X^+_{e^{'}}(1) = 1) - \mathbf{P}(X^-_{e^{'}}(1) = 1) \\ &\quad \leq (1 + \epsilon^{'})(1 + o(1))\partial_e \partial_{e'} H(X^+(0))\Psi^{'}(p^* + \epsilon)\varphi^{'}(p^*).\end{aligned} \qquad (\#\#)$$

Using the equations (#),(##) for sufficiently small $\epsilon, \epsilon^{'}$ we obtain the following inequality,

$$\begin{aligned}&\mathbb{E}d_H(X^+(1), X^-(1)) \leq \\ &1 - \binom{n}{2}^{-1} \left(1 - \sum_{e^{'} \neq e} \left(\mathbf{P}(X^+_{e^{'}}(1) = 1) - \mathbf{P}(X^-_{e^{'}}(1) = 1)\right)\right) \\ &\leq 1 - \binom{n}{2}^{-1} \left(1 - (1 + \epsilon^{''})(1 + o(1))\varphi^{'}(p^*)\right)\end{aligned}$$

The results hold since $\varphi^{'}(p^*) < 1$ (necessary condition). □

Next consider the monotone coupling starting from $X^+(0) = K_n$ and $X^-(0)$ empty. Define $\mathcal{A}_t = \{|r(G,e) - p^*| < \epsilon\}, \forall\, G, e\}$. Note from Corollary VI.6 for $t \geq cn^2$, $\mathbf{P}(\mathcal{A}_t) \leq 1 - e^{-\Omega(n)}$. Due to the monotonicity of Gibbs measure given $d = d_H(X^+(t), X^-(t))$ and the event $\mathcal{A}_t$ holds for $X^{\pm}(t)$, there exists a sequence

$$X^-(t) = X^0 \leq X^1 \leq \cdots \leq X^d = X^+(t),$$

with, 1) $d_H(X^{i+1}, X^i) = 1$ and 2) $\mathcal{A}_t$ holds for X^i, $\forall i$.

We have the following bound on the expected d_H at $(t+1)$,

$$\begin{aligned}&\mathbb{E}[d_H(X^+(t+1), X^-(t+1))] \\ &\quad \leq (1 - \delta n^{-2}) d_H(X^+(t), X^-(t)) + \binom{n}{2} e^{-\Omega(n)}.\end{aligned} \qquad (+)$$

The first part is due to path coupling on the given sequence and Lemma VI.9, whereas the second part is due to $d_H(X,Y) \leq \binom{n}{2}$ for any $X, Y \in \mathcal{G}_n$.

Iterating the equation (+) till $t^{'} = Cn^2$ and substituting $t > \frac{2+\epsilon}{\gamma} n^2 \log n$ we have $\mathbb{E}[d_H(X^+(t), X^-(t))] = o(1)$. Further application of Markov's inequality gives $\mathbf{P}(X^+(t)^-(t)) = o(1)$. Consequently, the mixing time is bounded by $\frac{2+\epsilon^{'}}{\gamma} n^2 \log n$. □

E. Proof of asymptotic independence of edges (Theorem IV.3) [1]

Fix an $\epsilon > 0$. Choose any subset of edges from the chosen k edges. Formally, let $S \subseteq [k]$, $x_S = \{x_{e_i} : i \in S\}$ and $x_{S^c} = \{x_{e_i} : i \in [k] \setminus S\}$. From inclusion and exclusion principal

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

We claim for sufficiently large n, $|\mathbf{P}(x_{S\cup T} = \mathbf{1}) - (p^*)^{|S \cup T|}| \leq \frac{\epsilon}{n^{|V|}}$. Therefore,

$$\left| \sum_{T \subseteq [k] \setminus S} (-1)^{|T|} \left(\mathbf{P}(x_{S \cup T} = \mathbf{1}) - (p^*)^{|S \cup T|}\right) \right| \leq \frac{\epsilon}{n^{|V|}}$$

The proof follows from the equality

$$\sum_{T \subseteq [k] \setminus S} (-1)^{|T|} (p^*)^{|S \cup T|} = (p^*)^{|S|} (1 - p^*)^{k - |S|}.$$

The above claim can be proved along the following line. For $\epsilon^{'}$ small enough, define

$$A = \left\{X : r_{max}(X) \leq p^* + \epsilon^{'}_{max}(X) \geq p^* - \epsilon^{'}\right\}.$$

Through some combinatorial arguments we can show for $\left|\mathbf{P}(x_T = \mathbf{1} | X \in A) - (p^*)^{|T|}\right| \leq \frac{\epsilon}{n^{|V|}}$. Next recalling for high temperature (single *mode*) we have $\mathbf{P}(A) = 1 - o(1)$ and so follows the claimed result. □

VII. Conclusion

The discussed work in [1] gives some basis for criticisms of the exponential random graph model. In high temperature, where fast mixing occurs, the model fails in capturing the clustering in real networks. Whereas, in low temperature, where it has the potential to model clustering, any local dynamics fail to give fast mixing. Having said that, we would like to point out for multi-modal distributions such as this, there are other MCMC techniques, e.g. Metroplis coupled MCMC, simulated tempering [8], [9], which have the potential to give overall fast mixing.

References

[1] S. Bhamidi, G. Bresler, and A. Sly, "Mixing time of exponential random graphs," in *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*. IEEE, 2008, pp. 803–812.

[2] O. Frank and D. Strauss, "Markov graphs," *Journal of the american Statistical association*, vol. 81, no. 395, pp. 832–842, 1986.

[3] S. Wasserman and P. Pattison, "Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp," *Psychometrika*, vol. 61, no. 3, pp. 401–425, 1996.

[4] T. A. Snijders, "Markov chain monte carlo estimation of exponential random graph models," *Journal of Social Structure*, vol. 3, no. 2, pp. 1–40, 2002.

[5] M. E. Newman, "Assortative mixing in networks," *Physical review letters*, vol. 89, no. 20, p. 208701, 2002.

[6] R. Bubley and M. Dyer, "Path coupling: A technique for proving rapid mixing in markov chains," in *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on*. IEEE, 1997, pp. 223–231.

[7] M. Dyer, A. Frieze, and M. Jerrum, "On counting independent sets in sparse graphs," *SIAM Journal on Computing*, vol. 31, no. 5, pp. 1527–1541, 2002.

[8] D. B. Woodard, S. C. Schmidler, and M. Huber, "Conditions for rapid mixing of parallel and simulated tempering on multimodal distributions," *The Annals of Applied Probability*, pp. 617–640, 2009.

[9] G. Altekar, S. Dwarkadas, J. P. Huelsenbeck, and F. Ronquist, "Parallel metropolis coupled markov chain monte carlo for bayesian phylogenetic inference," *Bioinformatics*, vol. 20, no. 3, pp. 407–415, 2004.