

# Understanding exponential random graph models

Mei Yin

Department of Mathematics  
University of Texas at Austin

January 25, 2012

Department of Mathematics, Michigan State University

The exponential family of random graphs is among the most widely-studied of network models. A host of analytical and numerical techniques have been developed in the past. We review recent developments in the study of exponential random graph models and concentrate on the phenomenon of phase transitions. We also present a new perspective: Any exponential random graph model could be alternatively viewed as a lattice gas model with a finite Banach space norm, and could be treated by cluster expansion methods in statistical mechanics.

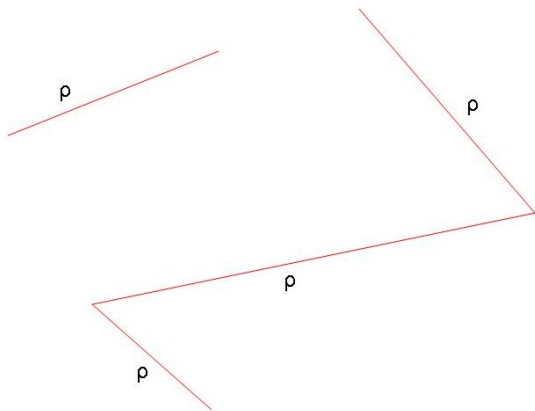
## Outline

- Introduction and Background
- Framework and Notation
- Recent Developments
- Alternative View
- Cluster Expansion

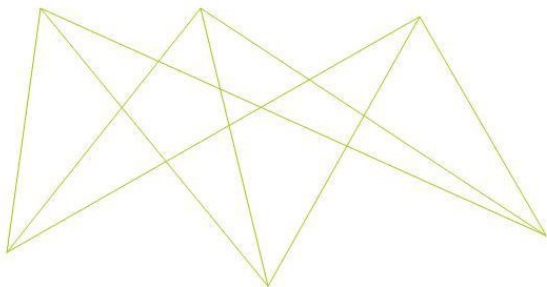
Pioneering work on the independent case: Erdős-Rényi graph  
 $G(n, \rho)$ ,

$$\mathbb{P}_n^\beta(G) = e^{\beta E(G) - \psi_n} = \rho^{E(G)} (1 - \rho)^{\binom{n}{2} - E(G)}.$$

Include edges independently with parameter  $\rho = e^\beta / (1 + e^\beta)$ .



Extremal Graph Theory (Turán)—Maximize number of edges without triangles. Unique solution: complete bipartite graph with equal parts.



More general statements ...

Exponential random graph: Dependence between the random edges is defined through certain finite subgraphs, in imitation of the use of potential energy to provide dependence between particle states in a grand canonical ensemble of statistical physics. By varying the activity parameters, one could analyze the extent to which specific values of the subgraph densities interfere with one another. Estimation can be based on construction of a Markov chain that has the exponential random graph model as equilibrium distribution. Large deviation principle comes into play.

- Holland and Leinhardt studied the directed case.
- Frank and Strauss related random graph edges to Markov random field.
- Häggström and Jonasson examined phase transition in the random triangle model.
- More developments: Wasserman and Faust, Snijders et al., Rinaldo et al.
- Recent survey: Fienberg, Introduction to papers on the modeling and analysis of network data I & II. arxiv: 1010.3882 & 1011.1717.

## Relevance to Gibbs measures:

- Ising model on complete graph: Curie-Weiss model. (Ellis and Newman, The statistics of Curie-Weiss models.)
- Ising model on sparse graph: No finite-dimensional structure. Distance between vertices. Phase transitions and coexistence phenomena are related to Gibbs measures on infinite trees. (Dembo and Montanari, Gibbs measures and phase transitions on sparse random graphs.)
- Ising model on lattice: Disordered limiting Gibbs state (with zero effective field) is pure up to the spin-glass critical temperature. (Bleher et al., On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice.)



Relevance to percolation: Obtained by independently removing vertices or edges from the graph.

Major difference: Random graphs have finite size, percolation systems are infinite.

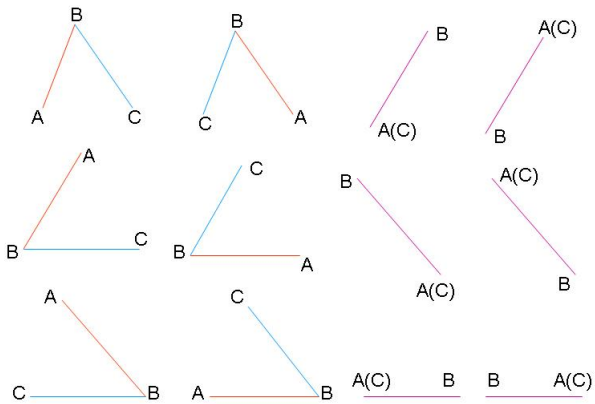
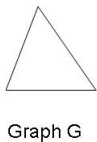
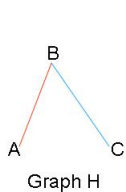
- van der Hofstad et al., First passage percolation on the random graph.
- Callaway et al., Network robustness and fragility: Percolation on random graphs.
- More: Chayes, Gandolfi, Spencer, ...

- Graph homomorphism  $\text{hom}(H, G)$  is an edge-preserving map. Examples:  $H$  triangle,  $G$  triangle,  $|\text{hom}(H, G)| = 6$ ;  $H$  2-star,  $G$  triangle,  $|\text{hom}(H, G)| = 12$ .
- Homomorphism density  $t(H, G) = \frac{|\text{hom}(H, G)|}{|V(G)|^{|V(H)|}$ .
- $\beta_1, \dots, \beta_k$  are  $k$  real parameters.  $H_1, \dots, H_k$  are finite simple graphs. Each  $H_i$  has  $m_i$  vertices ( $2 \leq m_i \leq m$ ) and  $p_i$  edges ( $1 \leq p_i \leq p$ ). In particular,  $H_1$  is the complete graph on 2 vertices (i.e., a single edge).
- $\mathcal{G}_n$  is the set of simple graphs  $G$  on  $n$  vertices. Probability for  $G \in \mathcal{G}_n$  is given by:

$$\mathbb{P}_n^{\{\beta_i\}}(G) = e^{n^2(\beta_1 t(H_1, G) + \dots + \beta_k t(H_k, G) - \psi_n)} := e^{n^2(T(G) - \psi_n)}.$$

- $\psi_n$  is the normalization constant:

$$\psi_n = \frac{1}{n^2} \log \sum_{G \in \mathcal{G}_n} e^{n^2 T(G)}.$$



$\psi = \lim \psi_n$  is crucial for carrying out maximum likelihood and Bayesian inference.

- Park and Newman tried the technique of mean-field approximations.
- Monte Carlo schemes: Geyer and Thompson (MCMLE), Gelman and Meng (bridge sampling), Kou et al. (equi-energy sampler)
- More approaches: Besag, Comets and Janžura, Chatterjee, Snijders

Consider the space  $\mathcal{W}$  of all symmetric measurable functions from  $[0, 1]^2$  into  $[0, 1]$ . For  $H \in \mathcal{G}_k$ , let

$$t(H, h) = \int_{[0,1]^k} \prod_{(i,j) \in E(H)} h(x_i, x_j) dx_1 \dots dx_k.$$

Lovász et al. developed graph limits (graphons): A sequence of graphs  $\{G_n\}_{n \geq 1}$  is said to converge to  $h$  if for every finite simple graph  $H$ ,

$$\lim t(H, G_n) = t(H, h).$$

Intuition: The interval  $[0, 1]$  represents a 'continuum' of vertices, and  $h(x, y)$  denotes the probability of putting an edge between  $x$  and  $y$ .

Example: Erdős-Rényi graph  $G(n, \rho)$ ,  $h(x, y) = \rho$ .

Example:

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

Then  $t(H, f^G) = t(H, G)$  for every finite simple graph  $H$ .

- Chatterjee and Diaconis gave the first rigorous proof of singular behavior in a specific exponential random graph model, the edge-triangle model. (Estimating and understanding exponential random graph models. arXiv: 1102.2650.)
- Related results: Bhamidi et al., Mixing time of exponential random graphs.
- Radin and Y derived the full phase diagram for a large family of 2-parameter exponential random graph models, each containing a first order transition curve  $\beta_2 = q(\beta_1)$  ending in a second order critical point (universality), qualitatively similar to the gas/liquid transition in equilibrium materials. (Phase transitions in exponential random graphs. arXiv: 1108.0649.)

Optimization problem: Suppose  $\beta_2, \dots, \beta_k$  are nonnegative. Then

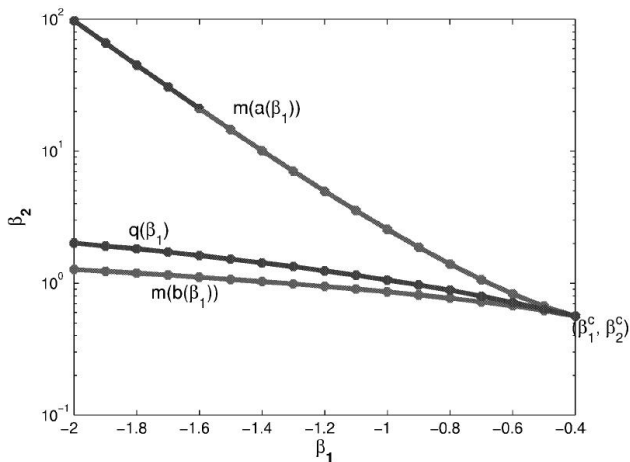
$$\begin{aligned} \psi = \\ \sup_{0 \leq u \leq 1} \left( \beta_1 u^{E(H_1)} + \dots + \beta_k u^{E(H_k)} - \frac{1}{2} u \log u - \frac{1}{2} (1-u) \log(1-u) \right) \\ := l(u^*). \quad (3.1) \end{aligned}$$

Behavior of  $G \in \mathcal{G}_n$ :

$$\min_{u^* \in U} \delta_{\square}(\tilde{G}, \tilde{u}^*) \rightarrow 0 \text{ in probability as } n \rightarrow \infty,$$

where  $U$  is the set of maximizers of (3.1).  $G$  behaves like the Erdős-Rényi graph  $G(n, u^*)$ .





The phase transition curve  $\beta_2 = q(\beta_1)$  in the  $(\beta_1, \beta_2)$  plane.  $H_1$  is a single edge and  $H_2$  has 3 edges.

- Chatterjee and Diaconis suggested that, quite generally, models with repulsion exhibit a transition qualitatively like the solid/fluid transition.  $\beta_2$  large negative.  $G$  looks like a complete  $(\chi(H) - 1)$ -equipartite graph.
- Aristoff and Radin showed that for  $\beta_2 < 0$  there is a curve  $\beta_2 = s(\beta_1)$  on which the exponential random graph model exhibits a phase transition. (Emergent structures in large networks. arXiv: 1110.1912.)

Improved proof by Y:  $\psi$  and  $I(u^*)$  coincide for  $\beta_2 > -\frac{2}{p(p-1)}$ . For fixed  $\beta_1$ ,

$$\lim_{\beta_2 \rightarrow -\infty} \psi = \frac{\chi(H) - 2}{2(\chi(H) - 1)} \log(1 + e^{2\beta_1}),$$

yet as  $I'(u^*) = 0$ ,

$$\lim_{\beta_2 \rightarrow -\infty} I(u^*) = \lim_{\beta_2 \rightarrow -\infty} \beta_2 (u^*)^p = 0.$$

Contradiction.

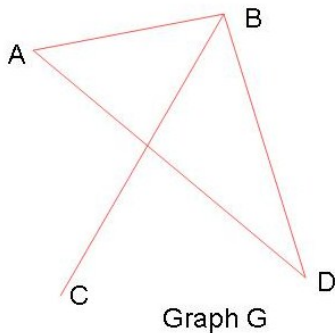
Fix  $H \in \mathcal{G}_m$ . Let  $G \in \mathcal{G}_n$ .

Proposition: The homomorphism density  $t(H, G)$  has a lattice gas representation  $\sum_X J(X)\sigma_X$ .

- $\sigma_{ij} = \sigma_{ji}$  is an element of the adjacency matrix of  $G$ .
- $X$  is any set of vertex pairs  $(i, j)$  of  $G$ .
- $\sigma_X = \prod_{(i,j) \in X} \sigma_{ij}$ .

Proof: Number of possible (connected) image shapes of  $H$  in  $G$  under the graph homomorphism is finite. Consider such an image shape  $Y$ . Denote the corresponding homomorphism density by  $t_Y(H, G)$ . Define  $J(X) = t_Y(H, G)$  for any  $X$  whose relative vertex positions are the same as in  $Y$ . This map becomes a homomorphism only when  $\sigma_X = 1$ , i.e., all corresponding edges between vertices in  $X$  exist.

Each edge,  $(A, B)$ ,  $(A, D)$ ,  $(B, C)$ ,  $(B, D)$ , carries weight  $2/4^3$ .



Each 2-star,  $(A, B, C)$ ,  $(A, B, D)$ ,  $(C, B, D)$ ,  $(B, A, D)$ ,  $(A, D, B)$ , carries weight  $2/4^3$ .

Proposition: Fix a vertex pair  $(i, j)$ , denote by  $t_{ij}(H, G)$  the part of the homomorphism density  $t(H, G)$  that depends on  $\sigma_{ij}$ , we have

$$t_{ij}(H, G) = \sum_{X:(i,j) \in X} J(X) \leq \frac{m(m-1)}{n^2}.$$

Note: Sharp bound. Example:  $H$  and  $G$  are both a single edge.

Proof: The image of  $V(H)$  in  $V(G)$  consists of vertices  $i$  and  $j$  of  $G$ . To count these homomorphisms, we regard such a mapping as consisting of two steps. Step 1: We choose the vertices of  $G$  that the vertices of  $H$  are mapped onto. Step 2: We check whether these vertex-maps are valid homomorphisms (i.e., edge-preserving).

Graphs in the same exponential random graph family correspond to equilibrium ensembles.

Hamiltonian:

$$H(\sigma) = -n^2 \sum_{i=1}^k \beta_i J_i(X) \sigma_X = - \sum_X K(X) \sigma_X$$

Note:  $K(X) = 0$  for  $|X| > p$ .

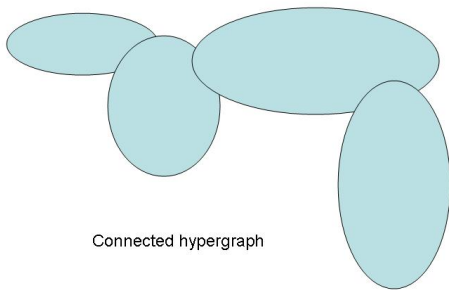
Banach space:

$$\|K\| = \sup_{(i,j)} \sum_{X:(i,j) \in X} |K(X)| \leq m(m-1) \sum_{i=1}^k |\beta_i|.$$

The limiting free energy (random graph model)  $\psi$  and the limiting free energy (lattice gas model)  $\phi$  are related by  $\psi = \frac{1}{2}(\log 2 + \phi)$ .



- Hypergraph  $\Gamma = (X, E)$ .
- $X$  is a set of sites,  $E$  (hyper-edge or link) is a set of nonempty subsets of  $X$ .
- Two links are connected if they overlap.
- Support of hypergraph:  $\cup \Gamma$ .
- Connected hypergraph  $\Gamma_c$ :  $\cup \Gamma$  is nonempty and cannot be partitioned into nonempty sets with no connected links.

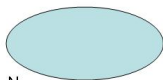
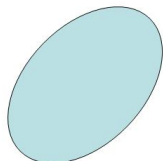
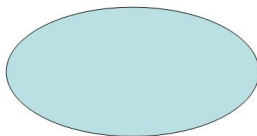


$W = \sum_{\sigma} e^{-H(\sigma)} = \sum_{\sigma} e^{\sum_X K(X)\sigma_X}$ . Let  $V$  be the set of all vertex pairs  $(i, j)$  of  $G \in \mathcal{G}_n$ .

Cluster representation for  $W$ :  $W = \sum_{\Delta} \prod_{N \in \Delta} w_N$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{N_1, \dots, N_n} \sum_G \prod_{(i,j) \in G} (-c(N_i, N_j)) w_{N_1} \cdots w_{N_n}.$$

- $\Delta$  is a set of disjoint subsets  $N$ 's of  $V$ .
- $w_N = \sum_{\cup \Gamma_c = N} \sum_{\sigma} \prod_{X \in \Gamma_c} (e^{K(X)\sigma_X} - 1)$ .
- $|w_N| \leq v_N = \sum_{\cup \Gamma_c = N} \prod_{X \in \Gamma_c} (e^{|K(X)|} - 1)$ .
- $c(N_i, N_j) = 1$  if  $N_i$  and  $N_j$  overlap; 0 otherwise.

 $N_1$  $N_2$  $N_3$ 

Typical term in  $W$ :  $w_{N_1} w_{N_2} w_{N_3}$

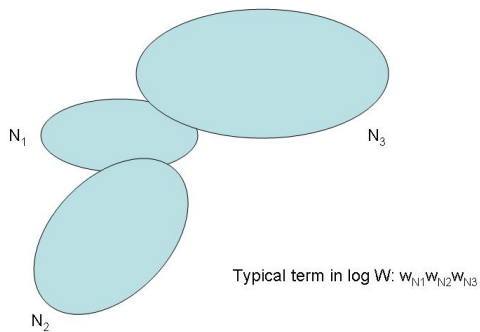
Cluster representation for  $\log W$ :

$$\log W = \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{N_1, \dots, N_n} C(N_1, \dots, N_n) w_{N_1} \cdots w_{N_n},$$

where

$$C(N_1, \dots, N_n) = \sum_{G_c} \prod_{(i,j) \in G_c} (-c(N_i, N_j)),$$

and  $G_c \in \mathcal{G}_n$  is a connected graph.



Kotecký-Preiss: Fix  $M > 1$ . Suppose that for each vertex pair  $(i, j)$ , we have

$$\sum_{N:(i,j) \in N} v_N M^{|N|} \leq \log M. \quad (5.1)$$

Then the pinned free energy has a convergent power series expansion:

$$\sum_{n=1}^{\infty} \frac{1}{n!} \sum_{N_1, \dots, N_n: \exists i N_i = N} |C(N_1, \dots, N_n)| |w_{N_1}| \cdots |w_{N_n}| \leq v_N M^{|N|}.$$

How is K-P applicable here? Consider the coupling constants  $K$  with the Banach space norm  $\|K\|$ . Suppose  $\sum_{i=1}^k |\beta_i|$  is small:

$$\|K\| \leq m(m-1) \sum_{i=1}^k |\beta_i| \leq \frac{\log M (p-1)^p}{2(Mp)^p (1 + (p-1) \log M)}.$$

Then (5.1) holds for every vertex pair  $(i, j)$ .

The maximal region of parameters  $\{\beta_i\}$  is obtained by setting

$$\log M = \frac{-p + \sqrt{5p^2 - 4p}}{2p(p-1)}.$$



Main result: Fix  $M > 1$ . Consider the coupling constants  $K$  with the Banach space norm  $\|K\|$ . Suppose  $\sum_{i=1}^k |\beta_i|$  is small. Then we have convergence of the cluster expansion for the limiting free energy  $\phi = \lim_{V \rightarrow \infty} \frac{1}{|V|} \log W$ .

$$\begin{aligned}
 |\log W| &\leq \sum_{N \subset V} \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{N_1, \dots, N_n: \exists i N_i = N} |C(N_1, \dots, N_n)| |w_{N_1}| \cdots |w_{N_n}| \\
 &\leq \sum_{N \subset V} v_N M^{|N|} \leq \sum_{(i,j) \in V} \sum_{N: (i,j) \in N} v_N M^{|N|} \leq |V| \log M.
 \end{aligned}$$

Proof:

$$\begin{aligned} \sum_{N:(i,j) \in N} v_N M^{|N|} &\leq \sum_{N:(i,j) \in N \cup \Gamma_c = N} \sum_{X \in \Gamma_c} M^{|N|} \prod_{X \in \Gamma_c} 2^{|K(X)|} \\ &\leq \sum_{\Gamma_c:(i,j) \in \Gamma_c} \prod_{X \in \Gamma_c} 2^{|K(X)|} M^{|X|}. \end{aligned}$$

Let

$$a_n(ij) = \sum_{(i,j) \in \Gamma_c: |\Gamma_c|=n} \prod_{X \in \Gamma_c} 2^{|K(X)|} M^{|X|}.$$

Then

$$\sum_{N:(i,j) \in N} v_N M^{|N|} \leq \sum_{n=1}^{\infty} \sup_{(i,j) \in V} a_n(ij) := \sum_{n=1}^{\infty} a_n.$$

To be continued ...

Lemma: Let  $a_n$  be the supremum over  $(i, j)$  of the contribution of connected hypergraphs with  $n$  links that are rooted at  $(i, j)$ . Then  $a_n$  satisfies the recursive bound

$$a_n \leq 2\|K\|M^p \sum_{k=0}^p \binom{p}{k} \sum_{a_{n_1}, \dots, a_{n_k} : n_1 + \dots + n_k + 1 = n} a_{n_1} \cdots a_{n_k} \quad (5.2)$$

for  $n \geq 1$ , where  $\binom{p}{k}$  is the binomial coefficient.

Lemma: Consider the coefficients  $\bar{a}_n$  that bound the contributions of connected and rooted hypergraphs with  $n$  links. Let  $w = \sum_{n=1}^{\infty} \bar{a}_n z^n$  be the generating function of these coefficients. The recursion relation (5.2) for the coefficients is equivalent to the formal power series generating function identity

$$w = 2\|K\|M^P z(1 + w)^P. \quad (5.3)$$

Lemma: If  $w$  is given as a function of  $z$  as a formal power series by the generating function identity (5.3), then this power series has a nonzero radius of convergence  $|z| \leq \frac{(p-1)^{p-1}}{2\|K\|(Mp)^p}$ .

Going back ...  $w = \sum_{n=1}^{\infty} \bar{a}_n z^n = 1/(p-1)$  corresponds to  $2\|K\|M^p z = (p-1)^{p-1}/p^p$ , this implies that for each  $n$ ,

$$\bar{a}_n \leq (2\|K\|(Mp)^p)^n (p-1)^{-(1+(p-1)n)}.$$

Gathering all the information we have obtained so far,

$$\begin{aligned} \sum_{N:(i,j) \in N} v_N M^{|N|} &\leq \sum_{n=1}^{\infty} (2\|K\|(Mp)^p)^n (p-1)^{-(1+(p-1)n)} \\ &= \frac{\frac{2\|K\|(Mp)^p}{(p-1)^p}}{1 - \frac{2\|K\|(Mp)^p}{(p-1)^{p-1}}} \leq \log M. \end{aligned}$$

# Thank You!

