

Structure and stability for sparse exponential random graphs

Conference on Random Matrix Theory and Numerical Linear Algebra

University of Washington, Seattle

24 June 2022

Nick Cook, Duke University

Based on joint works with [Amir Dembo](#) and [Huy Tuan Pham](#).

Exponential random graph models (ERGMs)

Let \mathcal{G}_n denote the set of simple graphs over $[n] = \{1, \dots, n\}$.

(We identify \mathcal{G}_n with $\{0, 1\}^{\binom{[n]}{2}}$.)

An ERGM is a probability measure on \mathcal{G}_n with mass function of the form

$$\frac{1}{Z_n(\alpha, \beta)} \exp(n^2 H(G; \beta) - \alpha e(G))$$

where $\alpha, \beta_1, \dots, \beta_m \in \mathbb{R}$ are the model parameters, $e(G) = \sum_{1 \leq i < j \leq n} G_{ij}$, and

$$H(G; \beta) = \sum_{k=1}^m \beta_k f_k(G)$$

for a fixed collection of graph statistics $f_1, \dots, f_m : \mathcal{G}_n \rightarrow \mathbb{R}$.

Ex. 1: Erdős–Rényi distribution. Taking $H \equiv 0$, $\alpha = \log \frac{1-p}{p}$ gives mass function $p^{e(G)}(1-p)^{\binom{n}{2}-e(G)}$.

Ex. 2: Edge-triangle model. $H(G; \beta) = \beta t(K_3, G)$, with

$$t(K_3, G) = \frac{1}{n^3} \sum_{i_1, i_2, i_3=1}^n G_{i_1, i_2} G_{i_2, i_3} G_{i_1, i_3}.$$

$$\frac{1}{Z_n(\alpha, \beta)} \exp \left(n^2 \sum_{k=1}^m \beta_k f_k(G) - \alpha e(G) \right)$$

- Introduced in the social sciences literature in the 80s–90s as parametric family of distributions for modeling social networks. [Frank & Strauss '86, Wasserman & Pattison '96].
- Want graphs with *transitivity*: friends of friends are more likely to be friends.
- The separable form of the Hamiltonian implies that the functions e, f_k are *sufficient statistics* for the model parameters α, β_k .
- Estimation of model parameters by MLE requires knowledge of the partition function $Z(\alpha, \beta)$, which is often done by sampling using local MCMC algorithms.

$$\frac{1}{Z_n(\alpha, \beta)} \exp \left(n^2 \sum_{k=1}^m \beta_k f_k(G) - \alpha e(G) \right)$$

Many problems in practice [Strauss '86, Snijders '02, Handcock '02, '03].

1. No transitivity!
2. **Degeneracy**: typical samples are either nearly empty or nearly full (edge density ~ 0 or ~ 1).
3. Slow convergence of sampling algorithms in some parameter regimes

$$\frac{1}{Z_n(\alpha, \beta)} \exp \left(n^2 \sum_{k=1}^m \beta_k t(F_k, G) - \alpha e(G) \right)$$

Bhamidi–Bressler–Sly '08: characterization of high/low-temperature regimes in “ferromagnetic” case $\beta_k > 0$.

- Low-temperature: Exponential convergence time for MCMC
- High-temperature: Polynomial convergence, but typical samples resemble Erdős–Rényi graphs (no transitivity!).

Chatterjee–Diaconis '12 (using **Chatterjee–Varadhan '11 LDP** for the Erdős–Rényi graph)

- Establish the naïve mean-field (NMF) variational approximation for the partition function.
- Ferromagnetic ERGMs are \approx mixture of Erdős–Rényi graphs.

Lubetzky–Zhao '12: Showed that for F Δ -regular, broken symmetry is restored for edge- F models of the form

$$\frac{1}{Z_n(\alpha, \beta)} \exp \left(n^2 \beta \mathbf{t}(F, G)^\gamma - \alpha \mathbf{e}(G) \right)$$

for Δ -regular F and a fractional power $\gamma \in (0, \frac{\Delta}{\mathbf{e}(F)})$.

Quantitative nonlinear LDT (Chatterjee–Dembo '14, Eldan '16, ...). Allowing n -dependent model parameters, density $p = o(1)$.

- Chatterjee–Dembo '14: Validity of NMF approximation.
- Eldan–Gross '18: ERGMs are approximately mixtures of stochastic block models, with barycenters close to critical points of NMF free energy.

Fix $h : \mathbb{R}_+^m \rightarrow \mathbb{R}$, and graphs F_1, \dots, F_m of max-degree $\Delta \geq 2$, and put

$$H(G) = h\left(\frac{t(F_1, G)}{p^{e(F_1)}}, \dots, \frac{t(F_m, G)}{p^{e(F_m)}}\right)$$

where

$$t(F, G) = \frac{1}{n^{v(F)}} \sum_{\phi: V(F) \rightarrow [n]} \prod_{\{u, v\} \in E(F)} G_{\phi(u), \phi(v)}$$

is the *homomorphism density* of F in G .

(Recall ex. $t(K_3, G) = \frac{1}{n^3} \sum_{i_1, i_2, i_3=1}^n G_{i_1, i_2} G_{i_2, i_3} G_{i_1, i_3}$.)

With large deviations rate parameter $r_{n,p} := n^2 \cdot p^\Delta \log(1/p)$ we define

$$\nu_{n,p}^H(\mathcal{E}) := \frac{\mathbb{E} e^{r_{n,p} H(\mathbf{G}_{n,p})} \mathbb{1}(\mathbf{G}_{n,p} \in \mathcal{E})}{\exp(\Lambda_{n,p}^H)}, \quad \mathcal{E} \subseteq \mathcal{G}_n.$$

with $\Lambda_{n,p}^H := \log \mathbb{E} \exp(r_{n,p} H(\mathbf{G}_{n,p}))$. (Tilt of Erdős–Rényi by $e^{r_{n,p} H(\cdot)}$.)

Naive mean-field (NMF) approximation and solution

$$\nu_{n,p}^H(\mathcal{E}) = \mathbb{E} \mathbb{1}(\mathbf{G}_{n,p} \in \mathcal{E}) \exp(r_{n,p} H(\mathbf{G}_{n,p}) - \Lambda_{n,p}^H)$$

Gibbs / Donsker–Varadhan variational principle:

$$\Lambda_{n,p}^H = \sup_{\mu} \{ r_{n,p} \mathbb{E}_{\mathbf{G} \sim \mu} H(\mathbf{G}) - D(\mu \| \mu_{n,p}) \}.$$

NMF approximation posits the sup is \approx attained on product distributions μ_Q , parametrized by set $\mathcal{Q}_n = [0, 1]^{\binom{[n]}{2}}$ of edge-weighted graphs $Q : \binom{[n]}{2} \rightarrow [0, 1]$.

$$\begin{aligned} \Lambda_{n,p}^H &\sim \sup_{Q \in \mathcal{Q}_n} \{ r_{n,p} \mathbb{E}_{\mathbf{G} \sim \mu_Q} H(\mathbf{G}) - D(\mu_Q \| \mu_{n,p}) \} \\ &\sim \sup_{Q \in \mathcal{Q}_n} \left\{ r_{n,p} H(Q) - \sum_{i < j} I_p(Q_{ij}) \right\} \quad =: \psi_{n,p}^H \end{aligned}$$

with $I_p(q) := D(\text{Ber}(q) \| \text{Ber}(p))$.

Main results I: NMF approximation

$$\Psi_{n,p}^H := \sup_{Q \in \mathcal{Q}_n} \left\{ r_{n,p} H(Q) - \sum_{i < j} I_p(Q_{ij}) \right\}.$$

Theorem (C.–Dembo '22)

Assume $n^{-1/(\Delta+1)} \ll p \leq 1$ and h is continuous and non-decreasing in each argument, with $h(x) = o_{\|x\| \rightarrow \infty}(\sum_{k=1}^m x_k^{\Delta/e(F_k)})$. Then $\Lambda_{n,p}^H = \Psi_{n,p}^H + o(r_{n,p})$.

When $p = o(1)$ we further reduce to a *2-dimensional* optimization problem. For any graph F there is an explicit function $T_F : [0, \infty)^2 \rightarrow [0, \infty)$ such that the following holds.

Theorem (C.–Dembo '22)

With hypotheses as in the previous result, assume further that $p = o(1)$. Then

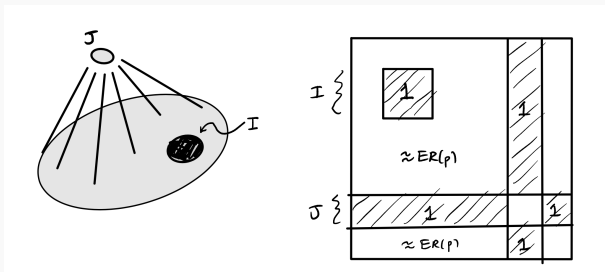
$$\frac{1}{r_{n,p}} \Lambda_{n,p}^H \rightarrow \psi(F, h) := \sup_{a, b \geq 0} \left\{ h(T_{F_1}(a, b), \dots, T_{F_m}(a, b)) - \frac{1}{2}a - b \right\}.$$

Main results II: Typical structure for sparse ERGMs

For $\varepsilon > 0$ and $I, J \subset [n]$, denote by $\mathcal{G}_n^{I,J}(\varepsilon)$ the set of graphs $G \in \mathcal{G}_n$ such that

$$\sum_{\{i,j\} \subset I} G_{ij} \geq (1 - \varepsilon) \frac{1}{2} |I|^2 \quad \text{and} \quad \sum_{i \in J, j \in J^c} G_{ij} \geq (1 - \varepsilon) |J| (n - |J|).$$

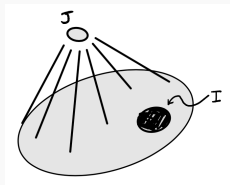
(G has an **almost-clique** at I and an **almost-hub** (biclique) at J).



$$\text{Set} \quad \mathcal{G}_n(a, b, \varepsilon) := \bigcup_{\substack{I, J \subset [n] \text{ disjoint} \\ |I| = \lfloor \sqrt{ap} \Delta^{1/2} n \rfloor, |J| = \lfloor bp \Delta n \rfloor}} \mathcal{G}_n^{I,J}(\varepsilon)$$

Main results II: Typical structure for sparse ERGMs

$$\mathcal{G}_n(a, b, \varepsilon) := \bigcup_{\substack{I, J \subset [n] \text{ disjoint} \\ |I| = \lfloor \sqrt{a} p^{\Delta/2} n \rfloor, |J| = \lfloor b p^{\Delta} n \rfloor}} \mathcal{G}_n^{I, J}(\varepsilon)$$



Theorem (C.–Dembo '22)

With $n^{-1/(\Delta+1)} \ll p \ll 1$ and \mathbf{h} as before, for any $\varepsilon > 0$ there exists $\eta = \eta(\mathbf{F}, \mathbf{h}, \varepsilon) > 0$ such that for $\mathbf{G} \sim \nu_{n,p}^{\mathbf{H}}$ and all n sufficiently large,

$$\mathbb{P}\left(\mathbf{G} \in \bigcup_{(a,b) \in \text{Opt}(\psi)} \mathcal{G}_n(a, b, \varepsilon)\right) \geq 1 - e^{-\eta n}$$

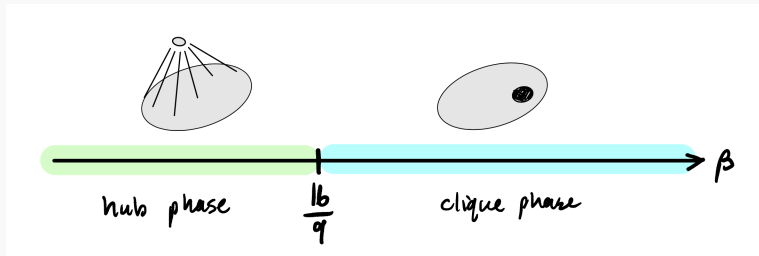
where $\text{Opt}(\psi) \subset \mathbb{R}_{\geq 0}^2$ is the set of optimizers for $\psi(\mathbf{F}, \mathbf{h})$.

Also have a stronger result in terms of spectral norm neighborhoods, in a narrower range of p .

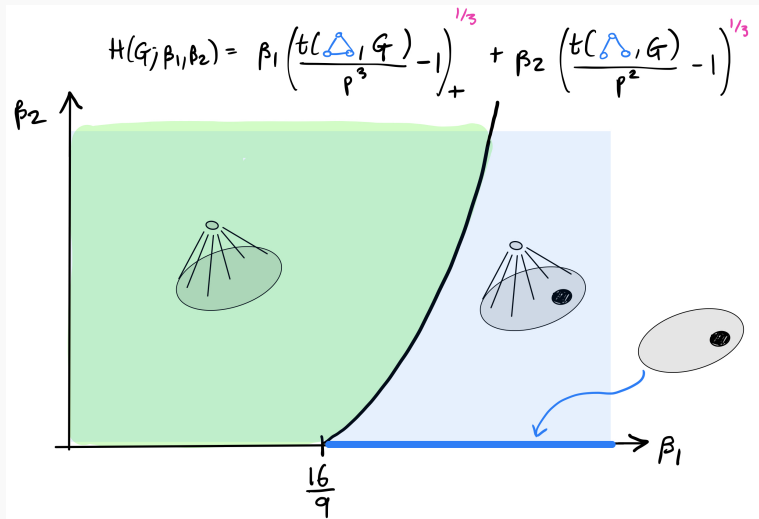
Example: (Tamed) edge-triangle model

Let $\mathbf{G} \sim \nu_{n,p}^H$ with $H(\mathbf{G}; \beta) = \beta \left(\frac{t(K_3, \mathbf{G})}{p^3} - 1 \right)_+^{1/3}$.

1. For fixed $\beta \in (0, \frac{16}{9})$, $\varepsilon > 0$, we have $\mathbf{G} \in \mathcal{G}_n(0, \frac{1}{3}\beta^{3/2}, \varepsilon)$ whp.
2. For fixed $\beta \in (\frac{16}{9}, \infty)$, $\varepsilon > 0$, we have $\mathbf{G} \in \mathcal{G}_n(\beta^2, 0, \varepsilon)$ whp.



Example: Edge- K_3 - P_3 model



Recall that a *cut matrix* is a rank-1 Boolean matrix $C = 1_S 1_T^T$. The following generalizes the classic Frieze–Kannan decomposition:

Decomposition Theorem (matrix case)

Assuming $np^{\Delta+1} \gg \log n$, for any fixed $\varepsilon > 0$ (small) and $K \geq 1$ (large), outside an exceptional set $\mathcal{E}_{\text{except}} \subset \{0, 1\}^{\binom{n}{2}}$ of measure $\mu_{n,p}(\mathcal{E}_{\text{except}}) \leq p^{cKn^2 p^\Delta}$, every adjacency matrix $A \in \{0, 1\}^{\binom{n}{2}}$ can be decomposed as

$$A = A_{\text{struct}} + A_{\text{rand}}$$

where A_{struct} is a linear combination of $O(K\varepsilon^{-2}p^{-\Delta})$ cut matrices C_i , with

$$\sum_{i=1}^k \|C_i\|_B \leq K\varepsilon^{-2}n^2 p^{\Delta-2} \quad \text{and} \quad \|A_{\text{rand}}\|_B^* \leq \varepsilon p.$$

From this we get a quantitative LDP.

Similar statement for r -tensors (under general B^* -norms).

- Result on typical structure for ERGMs is derived from a similar result for typical structure of $\mathbf{G}_{n,p}$ conditioned on joint upper tail events for $t(F_k, \mathbf{G}_{n,p})$, which in turn builds on previous works on large deviations for Erdős–Rényi graphs (Bhattacharya et. al '16, C.–Dembo '18, C.–Dembo–Pham '21).
- Main new ingredient: stability of optimizers in NMF problem $\Psi_{n,p}^H$.
- C.–Dembo–Pham '21 develops LDPs for random **hypergraphs**. ERHMs? Cf. Stasi et al '14.
- Related work on structure of random graphs picked uniformly under edge and F -count constraints. (Radin, Sadun et. al). Multipodality conjecture.
- Taming growth of Hamiltonian has “cured” the worst form of degeneracy, but maybe clique-hub graphs are also too degenerate for modeling social networks. Might get richer structure from degree constraints, antiferromagnetic models, other statistics $f_k(G)$, ...
- Also open: monotone decreasing h (e.g. $\beta_k < 0$). Need to analyze $\Psi_{n,p}^H$.