Asymptotics for exponential random graphs

Mei Yin[1]

Department of Mathematics, University of Denver

December 14, 2014

[1] Research supported under NSF grant DMS-1308333. Joint work with Richard Kenyon and with Lingjiong Zhu.

We begin with the definition of a k-parameter exponential family of random graphs (standard model).

Probability space: The set $\mathcal{G}_n$ of all simple graphs G_n on n vertices.
Probability mass function:

$$\mathbb{P}_n^{\beta}(G_n) = \exp\left(n^2(\beta_1 t(H_1, G_n) + ... + \beta_k t(H_k, G_n) - \psi_n^{\beta})\right).$$

- $\beta_1, ..., \beta_k$ are real parameters and $H_1, ..., H_k$ are pre-chosen finite simple graphs. Each H_i has vertex set $[k_i] = \{1, ..., k_i\}$ and edge set $E(H_i)$. By convention, we take H_1 to be a single edge.
- Graph homomorphism $\hom(H_i, G_n)$ is a random vertex map $V(H_i) \to V(G_n)$ that is edge-preserving. Homomorphism density $t(H_i, G_n) = \frac{|\hom(H_i, G_n)|}{|V(G_n)|^{|V(H_i)|}}$.
- Normalization constant:

$$\psi_n^{\beta} = \frac{1}{n^2} \log \sum_{G_n \in \mathcal{G}_n} \exp\left(n^2(\beta_1 t(H_1, G_n) + ... + \beta_k t(H_k, G_n))\right).$$

We begin with the definition of a k-parameter exponential family of random graphs (standard model).
Probability space: The set $\mathcal{G}_n$ of all simple graphs G_n on n vertices.
Probability mass function:

$$\mathbb{P}_n^{\beta}(G_n) = \exp\left(n^2(\beta_1 t(H_1, G_n) + ... + \beta_k t(H_k, G_n) - \psi_n^{\beta})\right).$$

- $\beta_1, ..., \beta_k$ are real parameters and $H_1, ..., H_k$ are pre-chosen finite simple graphs. Each H_i has vertex set $[k_i] = \{1, ..., k_i\}$ and edge set $E(H_i)$. By convention, we take H_1 to be a single edge.
- Graph homomorphism $\hom(H_i, G_n)$ is a random vertex map $V(H_i) \to V(G_n)$ that is edge-preserving. Homomorphism density $t(H_i, G_n) = \frac{|\hom(H_i, G_n)|}{|V(G_n)|^{|V(H_i)|}}$.
- Normalization constant:

$$\psi_n^{\beta} = \frac{1}{n^2} \log \sum_{G_n \in \mathcal{G}_n} \exp\left(n^2(\beta_1 t(H_1, G_n) + ... + \beta_k t(H_k, G_n))\right).$$

$\beta_i = 0$ for $i \geq 2$:

$$\begin{aligned} \mathbb{P}_n^{\beta}(G_n) &= \exp\left(n^2(\beta_1 t(H_1, G_n) - \psi_n^{\beta})\right) \\ &= \exp\left(2\beta_1 |E(G_n)| - n^2 \psi_n^{\beta}\right). \end{aligned}$$

Erdős-Rényi graph $G(n, \rho)$,

$$\mathbb{P}_n^{\rho}(G_n) = \rho^{|E(G_n)|}(1-\rho)^{\binom{n}{2} - |E(G_n)|}.$$

Include edges independently with parameter $\rho = e^{2\beta_1}/(1 + e^{2\beta_1})$.

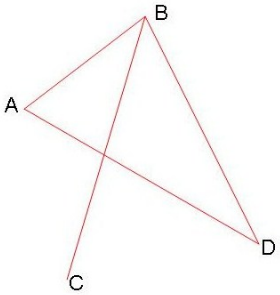
$$\exp(n^2 \psi_n^{\beta}) = \sum_{G_n \in \mathcal{G}_n} \exp\left(2\beta_1 |E(G_n)|\right) = \left(\frac{1}{1-\rho}\right)^{\binom{n}{2}}.$$

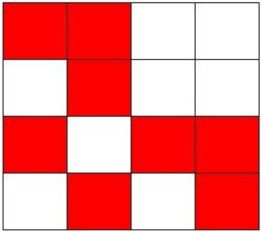
What happens with general β_i?
Problem: Graphs with different numbers of vertices belong to different probability spaces!
Solution: Theory of graph limits (graphons)! (Lovász and coauthors; earlier work of Aldous and Hoover)
Graphon space $\mathcal{W}$ is the space of all symmetric measurable functions $h(x, y)$ from $[0, 1]^2$ into $[0, 1]$. The interval $[0, 1]$ represents a 'continuum' of vertices, and $h(x, y)$ denotes the probability of putting an edge between x and y.

What happens with general β_i?
Problem: Graphs with different numbers of vertices belong to different probability spaces!
Solution: Theory of graph limits (graphons)! (Lovász and coauthors; earlier work of Aldous and Hoover)
Graphon space $\mathcal{W}$ is the space of all symmetric measurable functions $h(x, y)$ from $[0, 1]^2$ into $[0, 1]$. 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: Any $G_n \in \mathcal{G}_n$,

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





Why are we interested in exponential random graphs?

Dependence between the random edges is defined through certain finite subgraphs H_i, 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 β_i, one could analyze the extent to which specific values of the subgraph densities interfere with one another.

Why are we interested in exponential random graphs?
Dependence between the random edges is defined through certain finite subgraphs H_i, 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 β_i, one could analyze the extent to which specific values of the subgraph densities interfere with one another.

The normalization constant encodes useful information about the structure of the measure. By differentiating the normalization constant with respect to appropriate parameters, averages of various quantities of interest may be derived. Computation of the normalization constant is also essential in statistics because it is crucial for carrying out maximum likelihood estimates and Bayesian inference of unknown parameters.

Important contributors: Holland and Leinhardt; Frank and Strauss; Häggström and Jonasson; Wasserman and Faust; Snijders, Pattison, Robins, and Handcock; Rinaldo, Fienberg, and Zhou; Park and Newman; Krioukov; Chatterjee and Diaconis; Chatterjee and Dembo; Lubetzky and Zhao; Kenyon, Radin, Ren, and Sadun; Aristoff and Zhu; Hunter, Handcock, Butts, Goodreau, and Morris...

Survey: Fienberg, Introduction to papers on the modeling and analysis of network data I & II. arXiv: 1010.3882 & 1011.1717.

Large deviation and Concentration of measure:

$$\psi^{\beta} = \lim_{n\to\infty} \psi_n^{\beta} = \max_{h\in\mathcal{W}} \left(\beta_1 t(H_1, h) + ... + \beta_k t(H_k, h) - \int_{[0,1]^2} I(h) dxdy \right),$$

where:

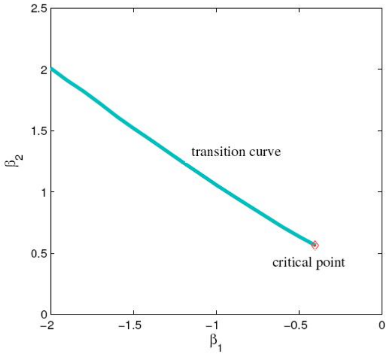
$$t(H_i, h) = \int_{[0,1]^{k_i}} \prod_{(i,j)\in E(H_i)} h(x_i, x_j) dx_1 ... dx_{k_i},$$

and $I : [0,1] \to \mathbb{R}$ is the function

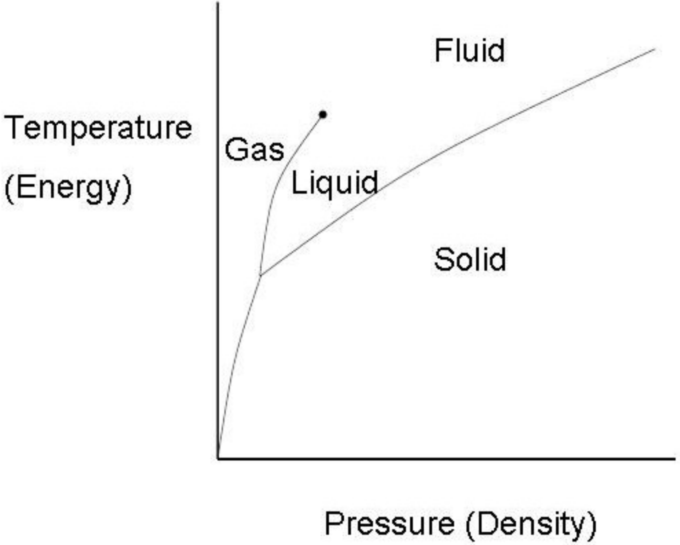
$$I(u) = \frac{1}{2} u \log u + \frac{1}{2}(1-u)\log(1-u).$$

Let F^* be the set of maximizers. G_n lies close to F^* with high probability for large n. (Chatterjee and Varadhan; Chatterjee and Diaconis)

Take H_1 a single edge and H_2 any finite simple graph. Then the parameter space $\{(\beta_1, \beta_2) : \beta_2 \geq 0\}$ consists of a single phase with a first-order phase transition across the indicated curve and a second-order phase transition at the critical point. (Radin and Y)



Graph drawn for H_2 a triangle. Critical point is $(\frac{1}{2}\log 2 - \frac{3}{4}, \frac{9}{16})$.


Temperature
(Energy)
Fluid
Gas
Liquid
Solid
Pressure (Density)

The standard exponential family of random graphs assumes no prior knowledge of the graph before sampling, but in many situations partial information of the graph is already known beforehand. What would be a typical random graph drawn from an exponential model subject to certain constraints?

Let $e \in [0,1]$ be a real parameter that signifies an "ideal" edge density. What happens if we only consider graphs whose edge density is close to e, say $|e(G_n) - e| < \alpha$?
(conditional) Probability mass function:

$$\mathbb{P}^{e,\beta}_{n,\alpha}(G_n) = \exp\left(n^2(\beta_1 t(H_1, G_n) + ... + \beta_k t(H_k, G_n) - \psi^{e,\beta}_{n,\alpha})\right) \cdot \\ \cdot \mathbb{1}_{|e(G_n)-e|<\alpha}.$$

(conditional) Normalization constant $\psi^{e,\beta}_{n,\alpha}$:

$$\psi^{e,\beta}_{n,\alpha} = \frac{1}{n^2} \log \sum_{G_n \in \mathcal{G}_n : |e(G_n)-e|<\alpha} \exp\left(n^2(\beta_1 t(H_1, G_n) + ... + \beta_k t(H_k, G_n))\right)$$

Large deviation and Concentration of measure:

$$\psi^{e,\beta} = \lim_{\alpha\to 0}\lim_{n\to\infty}\psi^{e,\beta}_{n,\alpha} = \beta_1 e +$$

$$\max_{h\in\mathcal{W}:e(h)=e}\left(\beta_2 t(H_2,h)+...+\beta_k t(H_k,h) - \int_{[0,1]^2} I(h)dxdy\right),$$

where:

$$e(h) = \int_{[0,1]^2} h(x,y)dxdy,$$

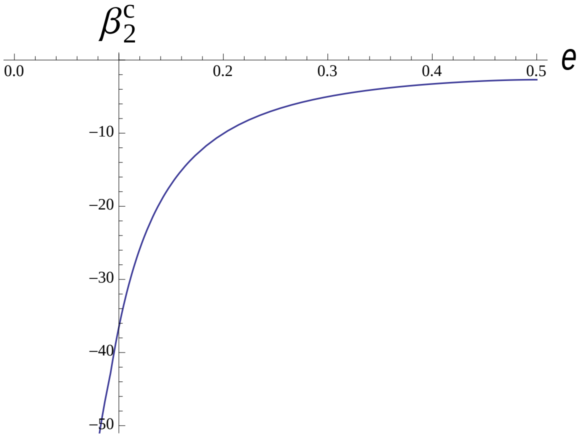
$$t(H_i,h) = \int_{[0,1]^{k_i}} \prod_{(i,j)\in E(H_i)} h(x_i,x_j)dx_1...dx_{k_i},$$

and $I : [0,1] \to \mathbb{R}$ is the function

$$I(u) = \frac{1}{2}u\log u + \frac{1}{2}(1-u)\log(1-u).$$

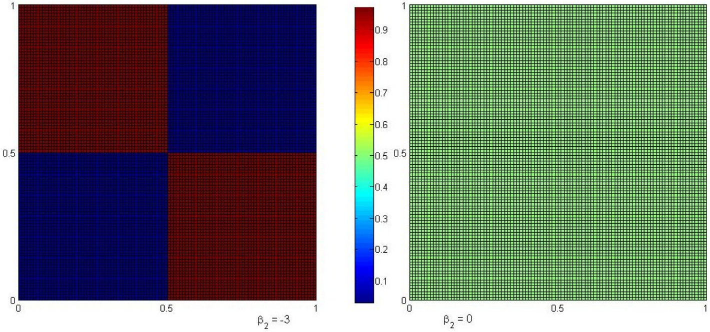
Let F^* be the set of maximizers. G_n lies close to F^* with high (conditional) probability for large n. (Kenyon and Y)

Take H_1 a single edge and H_2 a triangle. Fix the "ideal" edge density e. Let the edge parameter $\beta_1 = 0$ and the triangle parameter β_2 vary from 0 to $-\infty$. Then ψ^{e,β_2} loses its analyticity at at least one value of β_2. (Kenyon and Y)



The (conjectural) graph of non-analyticity point β_2^c as a function of e in the range $e \leq 1/2$.

On the special strip $e = \frac{1}{2}$, as β_2 decreases from 0 to $-\infty$, a typical graph G_n drawn from the constrained edge-triangle model jumps from being Erdős-Rényi to almost complete bipartite, skipping a large portion of the $e = \frac{1}{2}$ line.



All previous investigations have been centered on dense graphs (number of edges comparable to the square of number of vertices), but most networks data are sparse in the real world. What would be a typical random graph drawn from a sparse exponential model?

Let $\beta_i^{(n)} = \beta_i \alpha_n$ where $\alpha_n \to \infty$ as $n \to \infty$. For β_i negative this ensures that $\beta_i^{(n)} \to -\infty$ and translates to sparse graphs.
(sparse) Probability mass function:

$$\mathbb{P}_n^{\beta}(G_n) = \exp\left(n^2(\beta_1^{(n)} t(H_1, G_n) + ... + \beta_k^{(n)} t(H_k, G_n) - \alpha_n \psi_n^{\beta})\right).$$

(sparse) Normalization constant ψ_n^{β}:

$$\psi_n^{\beta} = \frac{1}{n^2 \alpha_n} \log \sum_{G_n \in \mathcal{G}_n} \exp\left(n^2(\beta_1^{(n)} t(H_1, G_n) + ... + \beta_k^{(n)} t(H_k, G_n))\right).$$

Large deviation and Concentration of measure:

$$\psi^{\beta} = \lim_{n\to\infty} \psi_n^{\beta} = \max_{h\in\mathcal{W}} \left(\beta_1 t(H_1, h) + ... + \beta_k t(H_k, h)\right),$$

where:

$$t(H_i, h) = \int_{[0,1]^{k_i}} \prod_{(i,j)\in E(H_i)} h(x_i, x_j) dx_1 ... dx_{k_i}.$$

Let F^* be the set of maximizers. G_n lies close to F^* with high probability for large n. (Y and Zhu)

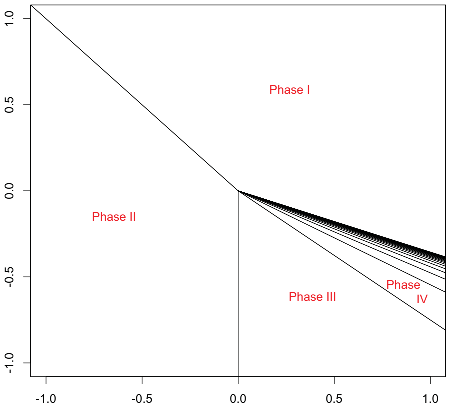
Let $X_{ij} = 1$ when there is an edge between vertex i and vertex j of G_n and let $X_{ij} = 0$ otherwise. Assume that $\lim_{n\to\infty} n^2 e^{2\alpha_n \beta_1} = 0$ and $\lim_{n\to\infty} \frac{\alpha_n}{n} = 0$. Then

$$\lim_{n\to\infty} \frac{\mathbb{P}_n^{\beta}(X_{1i} = 1)}{e^{2\alpha_n \beta_1}} = 1,$$

$$\lim_{n\to\infty} \frac{\mathbb{P}_n^{\beta}(X_{1i} = 1, X_{1j} = 1)}{e^{4\alpha_n \beta_1}} = 1, \quad i \neq j.$$

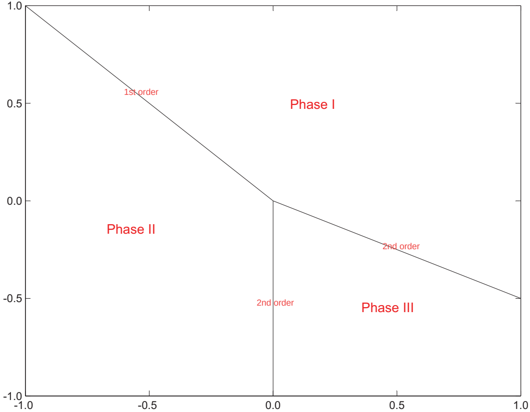
(Y and Zhu)

Take H_1 a single edge and H_2 a triangle. Then ψ^β undergoes countably many first-order phase transitions. (Y and Zhu)



Boundaries are given by $\{\beta_1 + \beta_2 = 0, \beta_1 < 0\}$, $\{\beta_1 = 0, \beta_2 < 0\}$, $\{\beta_1 = a_l \beta_2, \beta_1 > 0\}$, $\ell = 1, 2, ...$, and $\{\beta_1 = -3\beta_2, \beta_1 > 0\}$.

Take H_1 a single edge and H_2 a p-star. Then ψ^β exhibits both first- and second-order phase transitions. (Y and Zhu)



Graph drawn for H_2 a 2-star. Boundaries are given by $\{\beta_1 = 0, \beta_2 < 0\}$, $\{\beta_1 + \beta_2 = 0, \beta_2 > 0\}$, and $\{\beta_1 + 2\beta_2 = 0, \beta_2 < 0\}$.

Thank You!:)

After this conference, I realized that...

Thank You!:)

After this conference, I realized that...

Pitt is AWESOME even without the burgh!!