Statistical physics of exponential random graphs

Mei Yin[1]
メイ　イン

Department of Mathematics, University of Denver
デンバ – 大学　数学科

July 12, 2017

[1]Research supported under NSF grant DMS-1308333. This talk is based on joint work with multiple collaborators.

Large networks have become increasingly popular over the last decades, and their modeling and investigation have led to interesting and new ways to apply analytical and statistical methods. The introduction of exponential random graphs has aided in this pursuit, as they are able to capture a wide variety of common network tendencies by representing a complex global structure through a set of tractable local features. This talk will give an overview of phase transitions in large exponential random graphs. The main techniques that we use are variants of statistical mechanics but the exciting new theory of graph limits, which has rich ties to many parts of mathematics and beyond, also plays an important role in the interdisciplinary inquiry. Some open problems and conjectures will be presented.

Outline

- Standard exponential random graphs; Graph limit theory
- The edge-triangle model
- Constrained exponential random graphs; Statistical physics perspective
- Edge-weighted exponential random graphs; Universality

Erdős-Rényi graph $G(n, \rho)$: n vertices; include edges independently with probability ρ.

Empirical study of network structure shows that "transitivity is the outstanding feature that differentiates observed data from a pattern of random ties". Modeling transitivity (or lack thereof) in a way that makes statistical inference feasible however has proved to be rather difficult.

One direction is using exponential random graph models. They are particularly useful when one wants to construct models that resemble observed networks as closely as possible, but without going into detail of the specific process underlying network formation.

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 probability $\rho = e^{2\beta_1}/(1 + e^{2\beta_1})$.

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

What happens with general β_i?

Problem: Graphs with different numbers of vertices belong to different probability spaces!

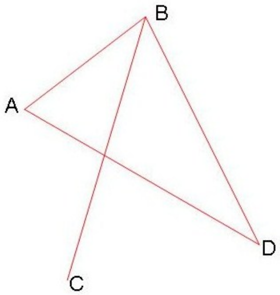
Solution: Theory of graph limits (graphons)! (Lovász and coauthors; earlier work of Aldous and Hoover)

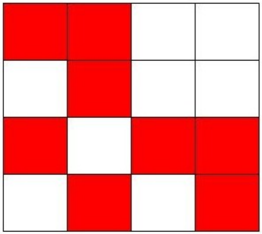
Graphon space $\mathcal{W}$ is the space of all symmetric measurable functions $h(x, y)$ from $[0, 1]^2$ into $[0, 1]$. The interval $[0, 1]$ represents a 'continuum' of vertices, and $h(x, y)$ denotes the probability of putting an edge between x and y.

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 $G_n \in \mathcal{G}_n$,

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





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

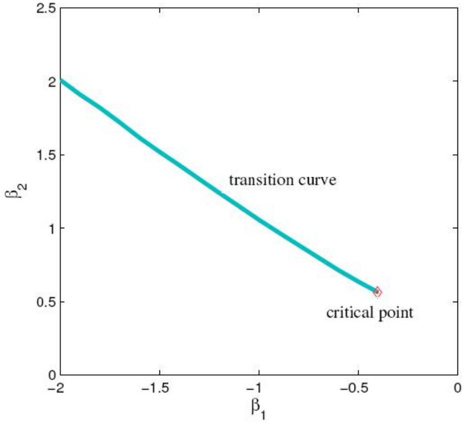
$$I(u) = \frac{1}{2} u \log u + \frac{1}{2}(1-u)\log(1-u).$$

Let F^* be the set of maximizers. G_n lies close to F^* with high probability for large n.
$\beta_2, ..., \beta_k \geq 0$: G_n behaves like the Erdős-Rényi graph $G(n, u^*)$, where $u^* \in [0, 1]$ maximizes

$$\beta_1 u + ... + \beta_k u^{|E(H_k)|} - \frac{1}{2} u \log u - \frac{1}{2}(1-u)\log(1-u).$$

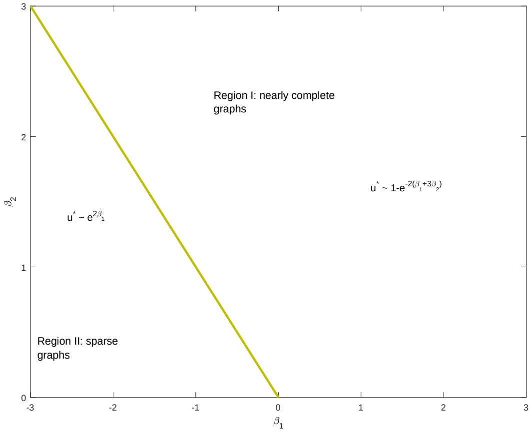
(Chatterjee and Varadhan; Chatterjee and Diaconis; Häggström and Jonasson; Bhamidi, Bresler, and Sly)

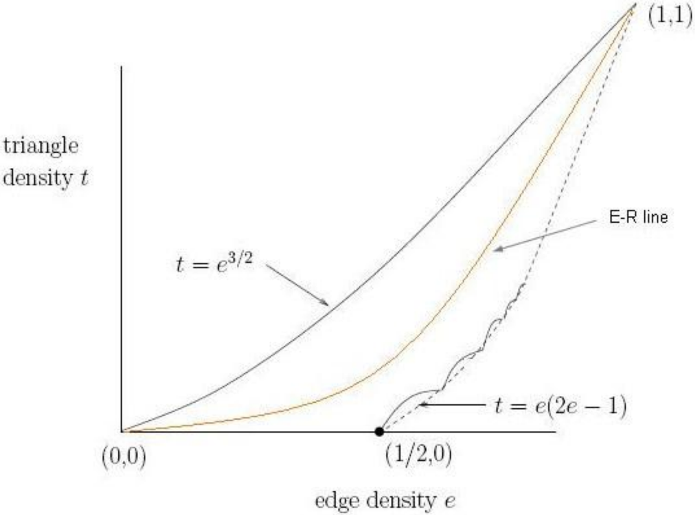
Take H_1 a single edge and H_2 a triangle. Fix the edge parameter β_1. Let the triangle parameter β_2 vary from 0 to ∞. Then ψ^{β_1,β_2} loses its analyticity at at most one value of β_2. (Radin and Y)



Critical point is $(\frac{1}{2}\log 2 - \frac{3}{4}, \frac{9}{16})$.

The line $\beta_1 = -\beta_2$ is of particular importance. The edge-triangle model transitions from an Erdős-Rényi type almost complete graph ($\beta_1 > -\beta_2$) to an Erdős-Rényi type almost empty graph ($\beta_1 \le -\beta_2$). (Y)

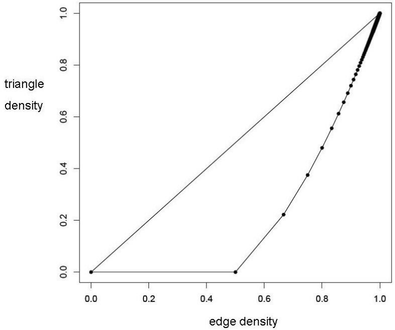


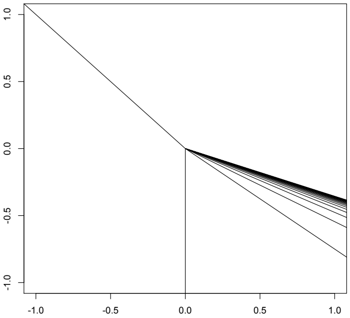


Feasible edge-triangle densities.

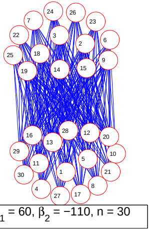
Upper bound: complete subgraph on $e^{1/2}n$ vertices.
Lower bound for $e \leq 1/2$: complete bipartite graph with $1-2e$ fraction of edges randomly deleted.
Lower bound for $e \geq 1/2$: complicated scallop curves where boundary points are complete multipartite graphs. (Razborov and others)

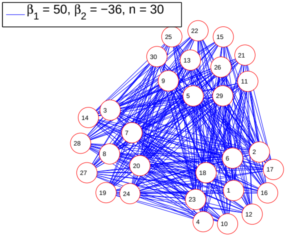
Take $\beta_1 = a\beta_2 + b$. Fix a and b. Let $n \to \infty$ and then let $\beta_2 \to -\infty$. G_n exhibits quantized behavior. (Y, Rinaldo, and Fadnavis; Mavi and Y; related work in Handcock; Rinaldo, Fienberg, and Zhou)

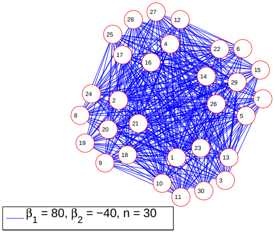




The infinite polytope.


β1 = 60, β2 = −110, n = 30


β1 = 50, β2 = −36, n = 30


β1 = 80, β2 = −40, n = 30

Picture the simple graph G_n as a realization of an Erdős-Rényi graph $G(n, .5)$. Let A_n be the adjacency matrix of G_n and $\text{tr}(\cdot)$ denote the trace of a matrix.
Alternate perspective for probability mass function:

$$\mathbb{P}_n^{\beta}(G_n) = \exp\left(\beta_1 \text{tr}(A_n^2) + \frac{\beta_2}{n}\text{tr}(A_n^3) - n^2\psi_n^{\beta}\right),$$

- $\beta_2 = 0$: $\psi_\infty^{\beta} = \log M(2\beta_1)/2$.
- $\beta_2 \to -\infty$: $\psi_\infty^{\beta} = \log M(2\beta_1)/4$.

$M(\theta) = (1 + \exp(\theta))/2$ is the moment generating function for Bernoulli $(.5)$ distribution.

The exponential family of random graphs have popular counterparts in statistical physics: a hierarchy of models ranging from the grand canonical ensemble, the canonical ensemble, to the microcanonical ensemble, with subgraph densities in place of particle and energy densities, and tuning parameters in place of temperature and chemical potentials.

The hierarchy

grand canonical ensemble $\longleftrightarrow$ exponential random graph
no prior knowledge of the graph is assumed
$\downarrow$
canonical ensemble $\longleftrightarrow$ constrained exponential random graph
partial information of the graph is given
$\downarrow$
microcanonical ensemble $\longleftrightarrow$ constrained graph
complete information of the graph is observed beforehand

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:

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

(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:

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

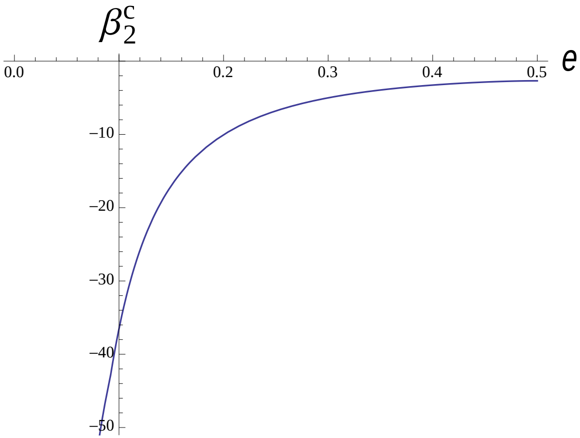
$$t(H_i,h)=\int_{[0,1]^{k_i}}\prod_{(i,j)\in E(H_i)} h(x_i,x_j)dx_1...dx_{k_i},$$

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

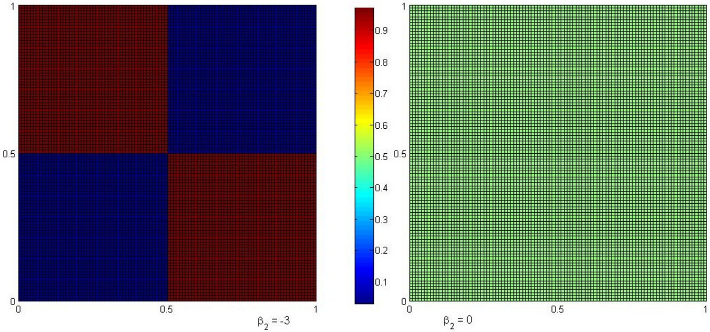
$$I(u)=\frac{1}{2}u\log u+\frac{1}{2}(1-u)\log(1-u).$$

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)



Special strip: Fix $e = \frac{1}{2}$. As β_2 decreases from 0 to $-\infty$, G_n jumps from Erdős-Rényi to almost complete bipartite, skipping a large portion of the $e = \frac{1}{2}$ line. (Kenyon and Y)



Simple graphs are such that the edge weights satisfy a Bernoulli (.5) distribution. Generalizations?

Probability space: The set $\mathcal{G}_n$ of all edge-weighted undirected graphs G_n on n vertices. Edge weights x_{ij} between vertices i and j are iid with a common distribution μ. This yields probability measure $\mathbb{P}_n$ and associated expectation $\mathbb{E}_n$ on $\mathcal{G}_n$.

Probability mass function:

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

Normalization constant ψ_n^{β}:

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

Simple graphs are such that the edge weights satisfy a Bernoulli (.5) distribution. Generalizations?

Probability space: The set $\mathcal{G}_n$ of all edge-weighted undirected graphs G_n on n vertices. Edge weights x_{ij} between vertices i and j are iid with a common distribution μ. This yields probability measure $\mathbb{P}_n$ and associated expectation $\mathbb{E}_n$ on $\mathcal{G}_n$.

Probability mass function:

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

Normalization constant ψ_n^β:

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

Take $\mu = \text{Unif}(0,1)$ as an example.
Large deviation and Concentration of measure:

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

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 Cramér's conjugate rate function

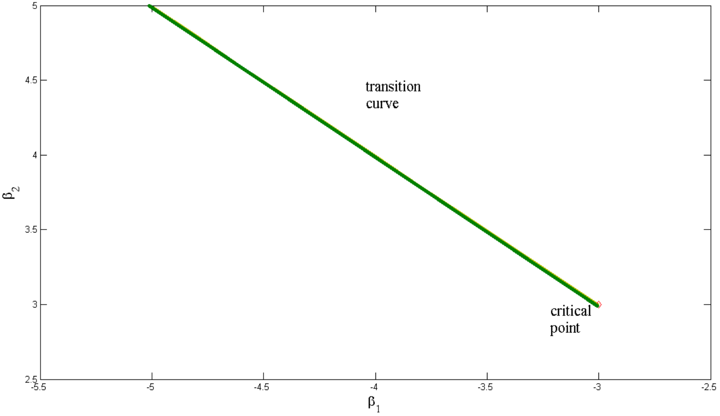
$$\begin{aligned} I(u) &= \sup_{\theta} \left(\theta u - \log \left(\int e^{\theta u} \mu(du) \right) \right) \\ &= \sup_{\theta} \left(\theta u - \log \frac{e^{\theta} - 1}{\theta} \right). \end{aligned}$$

Let F^* be the set of maximizers. G_n lies close to F^* with high probability for large n.
$\beta_2, ..., \beta_k \geq 0$: G_n behaves like the Erdős-Rényi graph $G(n, u^*)$, where $u^* \in [0, 1]$ maximizes

$$\beta_1 u + ... + \beta_k u^{|E(H_k)|} - \frac{1}{2} I(u).$$

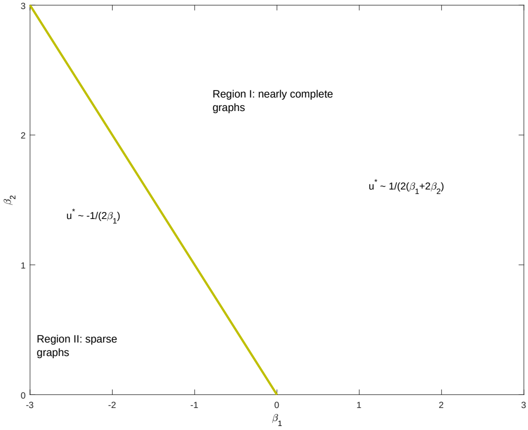
$I(u)$ does not admit closed-form expression; apply duality principle for Legendre transform. (Y)

Take H_1 a single edge and H_2 a 2-star. Fix the edge parameter β_1. Let the triangle parameter β_2 vary from 0 to ∞. Then ψ^{β_1,β_2} loses its analyticity at at most one value of β_2. (Y)



Critical point is $(-3, 3)$.

The line $\beta_1 = -\beta_2$ is of particular importance. The edge-triangle model transitions from an Erdős-Rényi type almost complete graph ($\beta_1 > -\beta_2$) to an Erdős-Rényi type almost empty graph ($\beta_1 \leq -\beta_2$). (Y)



Surprisingly (or not), the asymptotics of u^* heavily rely on the distribution μ; it is the asymptotics of θ^* (the dual of u^* following Legendre duality) that are universal. (DeMuse, Larcomb and Y)

- $\theta^* \approx 2(\beta_1 + \beta_2 |E(H_2)|)$ $(\beta_1 > -\beta_2)$
- $\theta^* \approx 2\beta_1$ $(\beta_1 \leq -\beta_2)$

The universal tendency of the phase transition curve is not intrinsic to symmetric distributions μ. Near degeneracy and universality are expected when the edge weights are not symmetrically distributed, except that the universal straight line gets shifted from $\beta_1 = -\beta_2$. Example: Take $\mu = \text{Bernoulli}(p)$. The universal line is asymptotically $\beta_1 = -\beta_2 - \log(p/(1-p))/2$, which corresponds to an upward shift from $\beta_1 = -\beta_2$ for $p < 1/2$ and a downward shift for $p > 1/2$.

Surprisingly (or not), the asymptotics of u^* heavily rely on the distribution μ; it is the asymptotics of θ^* (the dual of u^* following Legendre duality) that are universal. (DeMuse, Larcomb and Y)

- $\theta^* \approx 2(\beta_1 + \beta_2 |E(H_2)|)$ $(\beta_1 > -\beta_2)$
- $\theta^* \approx 2\beta_1$ $(\beta_1 \leq -\beta_2)$

The universal tendency of the phase transition curve is not intrinsic to symmetric distributions μ. Near degeneracy and universality are expected when the edge weights are not symmetrically distributed, except that the universal straight line gets shifted from $\beta_1 = -\beta_2$. Example: Take $\mu = \text{Bernoulli}(p)$. The universal line is asymptotically $\beta_1 = -\beta_2 - \log(p/(1-p))/2$, which corresponds to an upward shift from $\beta_1 = -\beta_2$ for $p < 1/2$ and a downward shift for $p > 1/2$.

Thank You!:)

ありがとうございました!:)

I have two close relatives who graduated from Waseda, both during the 1930's, but this is the first time that I'm here! And I loved everything about it!:) Thank you very much for the wonderful opportunity!