Weighted exponential random graph models: scope and large network limits

Suman Chakraborty

University of North Carolina, Chapel hill

sumanc@live.unc.edu

August 21, 2017

Plan for the talk

Based on a joint work with Bhamidi, Cranmer, Desmarais ’17.

- Why weighted network?
- Exponential models for simple(binary) and weighted network.
- Main tool: Dense graph limits.
- (Graph) Limiting results on (G)ERGM and applications.
- A special case: Normal distribution.
- Discussions.

Why weighted network?

What is a simple(binary) network?
Graph with vertex set $V := [n]$ and edge set E, where $(i,j) \in E$ if nodes i and j are connected for $i,j \in n$. The $n \times n$ adjacency matrix of (V,E) is the matrix X with elements

$$\begin{aligned} x_{ij} &= 1 \quad \textit{if} \quad (i,j) \in E, \\ &= 0 \quad \textit{if} \quad (i,j) \in E^c. \end{aligned}$$

- Binary network captures the information about the connectivity structure of the vertex set.
- It does not captures information about the **strength** of the connection.

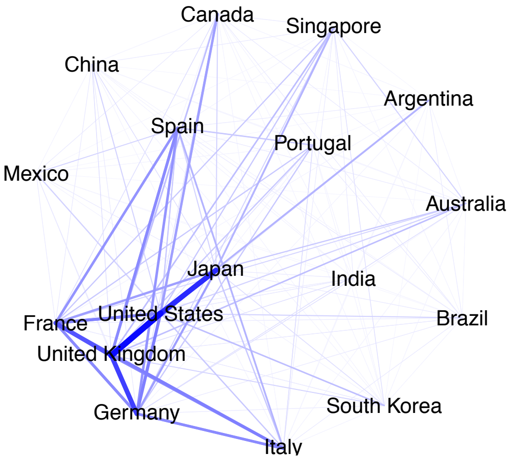

Figure: International Lending Network 2005.(The plot is taken from Wilson, Denny, Bhamidi, Cranmer, Desmarais '16)

Motivating Examples

- A binary network would require thresholding and thus loss of information.
- Instead of using binary values we associate an weight to each pair of nodes, the adjacency matrix becomes $x_{ij} = \{\text{Weight of the edge } (i,j)\}$.
- The weights can be any real number.

Statistical Motivation

Studies by Park and Newman ('04a); Wasserman and Pattison ('96) Snijders, Pattison, Robins and Handcock ('06); Fienberg ('10)...suggest,

- High variance in the "popularity" of nodes implies high values of $\sum_{i,j,k} x_{ij}x_{ik}$. (This is homomorphism number of two-star)
- High transitivity indicates high values of $\sum_{i,j,k} x_{ij}x_{jk}x_{ki}$. (This is homomorphism number of triangles)
- Higher values of other motif counts in many applications.

Exponential Random Graph Models(ERGM) were used to incorporate the above features.

An Examples of ERGMs

Following is a model on the space of all simple graphs(binary and undirected) on n nodes,

$$p_{\beta_1,\beta_2}(G) \propto \exp\left(2\beta_1 E + 6\frac{\beta_2}{n} TR\right),$$

$E =$ Number of edges in G and $TR =$ Number of triangles in G.
Intuition,

- What happens if $\beta_2 = 0$?
- What happens if $\beta_2 > 0$?
- Why the scaling n is required for the triangle?
- Why the constants 2 and 6 are there?

Will come back to the Model after a (very)short overview of graph limits.

Brief Overview Of Graph Limits

Map the $n \times n$ adjacency matrix $X = \{x_{ij}\}_{1 \leq i,j \leq n}$ with $x_{ij} = x_{ji}$ for all $i, j \in [n]$ to a symmetric kernel:

$$k(x, y) = \sum_{i,j=1}^{n} x_{ij} \mathbf{1}_{J_i^n}(x) \mathbf{1}_{J_j^n}(y),$$

where $J_1^n = [0, \frac{1}{n}]$ and for $i = 2, \ldots, n$, J_i^n is the interval $(\frac{i-1}{n}, \frac{i}{n}]$.
$\mathcal{K}$: Space of symmetric measurable functions from $[0,1] \times [0,1] \to \mathbb{R}$. The cut distance in the space $\mathcal{K}$ is defined as follows,

$$d(k_1, k_2) = \sup_{A,B \subset [0,1]} \left| \int_{A \times B} (k_1(x,y) - k_2(x,y))\, dx\, dy \right|,$$

where A and B are Borel subsets of $[0, 1]$.

- Σ : space of all measure preserving bijections (with respect to the Lebesgue measure) $\sigma : [0,1] \to [0,1]$.
- $k_1, k_2 \in \mathcal{K}$, say that $k_1 \sim k_2$ if,

$$k_1(x,y) = \sigma k_2(x,y) := k_2(\sigma x, \sigma y), \qquad \text{a.e. } x, y, \qquad \text{for some } \sigma \in \Sigma.$$

- Denote the orbit $\{\sigma k : \sigma \in \Sigma\}$ by $\tilde{k}$. Write $\tilde{\mathcal{K}} := \mathcal{K} \backslash \sim$ for the quotient space under the relation $\sim$ on $\mathcal{K}$ and τ for the natural map from $k \to \tilde{k}$.
- d is invariant under σ, so a natural distance δ on $\tilde{\mathcal{K}}$ is:

$$\delta(\tilde{k_1}, \tilde{k_2}) = \inf_{\sigma} d(\sigma k_1, k_2) = \inf_{\sigma} d(k_1, \sigma k_2) = \inf_{\sigma_1, \sigma_2} d(\sigma_1 k_1, \sigma_2 k_2).$$

Two results and an important representation

Lovász and coauthors proved many important results about this metric space.

- $(\tilde{\mathcal{K}}^t, \delta)$ is a compact metric space, where $\tilde{\mathcal{K}}^t = \{\tilde{k} \in \tilde{\mathcal{K}} : |\tilde{k}| \leq t\}$ for $t \in \mathbb{R}$.

Why is this metric useful?

- It makes many important functionals continuous. e.g, **homomorphism** density, normalized spectra and many more.

How one represents **homomorphism** density of a graph F into a Kernel k?

$$t(F, k) = \int_{[0,1]^{|V(F)|}} \prod_{(i,j) \in E(F)} k(x_i, x_j) \prod_{i \in V(F)} dx_i$$

Note: $t(F, k)$ is invariant under measure preserving transformation.

Two Examples

E be an edge and G be a graph on n vertices and mapped into the kernel k_G. Then

$$t(E, G) = t(E, k_G) = \int_{[0,1]} k_G(x_1, x_2)\, dx_1\, dx_2 = 2 \text{ no. of edges in } G.$$

TR be a triangle, then,

$$t(TR, G) = t(TR, k_G) = \int_{[0,1]^3} k_G(x_1, x_2) k_G(x_2, x_3) k_G(x_3, x_1)\, dx_1\, dx_2\, dx_3$$
$$= 6 \text{ no. of triangles in } G.$$

The GERGM

- With each edge $\{i, j\}$, assign an i.i.d probability distribution $q_{ij} (= q_{ji})$ for $1 \leq i < j \leq n$.
- For the diagonal, $q_{ii} = \delta_0$ be the unit mass at zero, for $i = 1, \ldots, n$ independent of the remaining edges.
- Write Q_n for the induced measure on $\mathcal{K}$ via the mapping of the graph into the kernel on $[0, 1] \times [0, 1]$ and $\tilde{Q}_n$ for the corresponding push-forward measure on $\tilde{\mathcal{K}}$. Call $\tilde{Q}_n$ the **base measure**.

The generalized exponential random graph model is a probability measure $\tilde{R}_n$ on $\tilde{\mathcal{K}}$ defined via **tilting** $\tilde{Q}_n$ using a given function T.

$$d\tilde{R}_n(\tilde{k}) = \exp\{n^2(T(\tilde{k}) - \psi_n)\} \, d\tilde{Q}_n(\tilde{k}), \qquad \tilde{k} \in \tilde{\mathcal{K}}.$$

where ψ_n is the normalizing constant.

Literature and Challenges

Desmarais and Cranmer '12, Krivitsky '12, Wilson, Denny, Bhamidi, Cranmer, Desmarais '16 used weighted exponential random graphs with various choices of base measures and statistics T. Following are some of the main challenges:

- Estimating the normalizing constant.
- Understand large network limits.
- "Degeneracy" and "No Degeneracy" phenomenon.
- Choice of base measures and statistics T.

Normalizing Constant

One main challenge is to estimate the normalizing constant of GERGM.
Theorem(Bhamidi, C, Cranmer, Desmarais '17). *Under some assumptions on the base measure and the statistic T, the limiting normalizing constant is given by,*

$$\psi = \lim_{n\to\infty} \psi_n = \lim_{l\to\infty} \sup_{\tilde{k}\in\tilde{\mathcal{K}}^l} (T(\tilde{k}) - I(\tilde{k})),$$

where

$$I(k) = \frac{1}{2}\iint_{[0,1]\times[0,1]} h(k(x,y))\,dx\,dy$$

with $h(x) := \sup_\theta [\theta x - \ln M(\theta)]$, and $M(\theta)$ *is the moment generating function of* q.

Assumptions

(C1: Finiteness) Suppose for each fixed $t > 0$, T is a bounded continuous function in cut metric when restricted to $\mathcal{K}^t$ and further satisfies

$$\int_{\tilde{\mathcal{K}}} \exp(n^2 T(\tilde{k}))\, d\tilde{Q}_n(\tilde{k}) < \infty,$$

(C2: Exponential tightness)

$$\limsup_{l\to\infty} \limsup_{n\to\infty} \frac{1}{n^2} \ln \int_{\{T(\tilde{k}) - T(f_l(\tilde{k})) \geq \varepsilon\}} e^{n^2 T(\tilde{k})}\, d\tilde{Q}_n(\tilde{k}) = -\infty.$$

where

$$\begin{aligned} f_l(q) &= q \text{ if } |q| \leq l, \\ &= l \text{ if } q \geq l, \\ &= -l \text{ if } q \leq -l. \end{aligned}$$

Proof Technique

The proof involves the following steps:

- Noting $\psi_n = \frac{1}{n^2} \ln \int_{\tilde{\mathcal{K}}} e^{n^2 T(\tilde{k})} \, d\tilde{Q}_n(\tilde{k})$.
- Truncate the integral using operator f_l.
- Show that the $\{T(\tilde{k}) - T(f_l(\tilde{k})) \geq \varepsilon\}$ does not contribute anything at n^2 scale using our assumptions.
- Using Large deviations result by Chatterjee and Varadhan '11 together with continuity of T and compactness of $\tilde{\mathcal{K}}^l$.

A Law of Large numbers

Theorem(Bhamidi, C, Cranmer, Desmarais '17). Let $\tilde{F}^*$ *be the set of maximizers of* $T(\tilde{k}) - I(\tilde{k})$. *If together with* **(C1)** *and* **(C2)** *we assume all elements in* $\tilde{F}^*$ *are absolutely bounded, then for any* $\eta > 0$ *there exist constants* $C, \gamma > 0$ *such that,*

$$R_n(\delta(\tilde{k}, \tilde{F}^*) \geq \eta) \leq Ce^{-n^2\gamma},$$

for all $n \geq 1$.

Observations from GERGM are concentrated around the set $\tilde{F}^*$.

Exact solutions for some popular GERGM

H_1 be a graph with two vertices and a single edge joining these vertices and H_i's are graphs with at least two edges for $2 \leq i \leq s$. We will consider the following statistics,

$$T(k) = \sum_{i=1}^{s} \beta_i t(H_i, k),$$

Theorem(Bhamidi, C, Cranmer, Desmarais '17). *Under assumptions* **(C1)** *and* **(C2)**, *where $\beta_2, \ldots, \beta_s$ are non negative real numbers. Also suppose either the kernel k is non-negative or $e(H_i)$'s are even positive integers for all $2 \leq i \leq s$. Then the value of the normalizing constant is given by*

$$\lim_{n \to \infty} \psi_n = \sup_{u \in \mathbb{R}} \left(\sum_{i=1}^{s} \beta_i u^{e(H_i)} - I(u) \right)$$

Concentration of the model

Last theorem made evaluating the limiting normalizing constant a simple optimization problem. The following theorem gives how a typical observation would look like.

Theorem(Bhamidi, C, Cranmer, Desmarais '17). *In addition to the usual assumptions assume that*

$$\lim_{|u|\to\infty} \sum_{i=1}^{s} \beta_i u^{e(H_i)} - I(u) = -\infty.$$

Let K be the set of maximizers of the function $g(\cdot)$ defined via $g(u) := \sum_{i=1}^{s} \beta_i u^{e(H_i)} - I(u)$. Then K has finitely many elements and

$$\min_{u\in K} \delta(\tilde{k}_n, \tilde{k}^u) \to 0,$$

as $n \to \infty$, almost surely, where $\tilde{k}^u$ are the constant kernel equal to u on $[0,1]\times[0,1]$.

Comments

- The last two theorems enable us to evaluate the normalizing constant easily.
- If for some β, $g(u) := \sum_{i=1}^{s} \beta_i u^{e(H_i)} - I(u)$ has an unique maximizer then it is sometimes called **"high temperature"** regime.
- The concentration result shows that under the high temperature regime the model is essentially indistinguishable from an independent random graph with edge probability is a function of $\beta := (\beta_1, \ldots, \beta_s)$.
- It does not say anything when $\beta_i < 0$ for some $i \geq 2$.

A Special Case

Consider the model of the form where H_1 is a single edge as before and H_j's are j-stars for all $2 \leq j \leq s$.

Theorem(Bhamidi, C, Cranmer, Desmarais '17). *Under usual assumptions the value of the normalizing constant is given by*

$$\lim_{n\to\infty} \psi_n = \sup_{u\in\mathbb{R}} \left(\sum_{i=1}^{s} \beta_i u^{e(H_i)} - I(u) \right)$$

Let K be the set of maximizers of the function $g(\cdot)$ defined via $g(u) := \sum_{i=1}^{s} \beta_i u^{e(H_i)} - I(u)$. Assume that,

$$\lim_{|u|\to\infty} \sum_{i=1}^{s} \beta_i u^{e(H_i)} - I(u) = -\infty.$$

Then K has finitely many elements and

$$\min_{u\in K} \delta(\tilde{k}_n, \tilde{k}^u) \to 0.$$

as $n \to \infty$, almost surely.

Degeneracy Phenomenon

Consider the GERGM model,

$$T(k) = \beta_1 t(E, k) + \beta_2 t(TR, k),$$

with base measure Bernoulli(1/2).

- Analysis by Handcock '03 suggested if β_1 is large negative number and β_2 varies then the edge density in the resulting graph goes from **very small(close to zero) to very large(close to one) skipping all intermediate values**.
- Park and Newman '04 suggested this phenomenon for Edge, Two-Star ERGM.
- Chatterjee and Diaconis '13 gave first rigorous proof of this phenomenon for the Edge-Triangle ERGM.
- Radin and Yin '13 gave detailed analysis of this phenomenon for a large class of ERGM models.

Degeneracy Phenomenon

- This is problematic in practice.
- How to detect these “unwanted” regions?
- What happens in GERGM?
- A detailed simulation study by Wilson, Denny, Bhamidi, Cranmer, Desmarais ’16 suggested: Edge-Two-star model with base measure truncated normal distribution does not suffer from degeneracy.

Normal Distribution and Degeneracy Phenomenon

Theorem(Bhamidi, C, Cranmer, Desmarais '17). *Consider the model with* $T(k) = \beta_1 t(H_1, k) + \beta_2 t(H_2, k)$ *with* H_1 *is an edge and* H_2 *is two-star and standard normal distribution as the base measure.*

$$\psi_n = \frac{1}{\sqrt{1 - \frac{4\beta_2(n-1)}{n}}} \exp\left(\frac{\beta_1^2 n(n-1)}{1 - 4\beta_2 \frac{(n-1)}{n}}\right)\left(1 - \frac{2\beta_2(n-2)}{n}\right)^{-\frac{(n-1)}{2}}.$$

whenever $n \geq 3$ *and* $\beta_2 < \frac{n}{4(n-1)}$. *In particular,*

$$\lim_{n\to\infty} \frac{1}{n^2} \ln \psi_n = \left(\frac{\beta_1^2}{1 - 4\beta_2}\right).$$

Remark: The model **does not** suffer from degeneracy.

Summary

- Derived the limiting normalizing constant.
- Understood large network limit of GERGM.
- Derived formula for limiting Normalizing constant for large class of models.
- Proved no degeneracy for edge-two star model with standard normal base measure.

THANK YOU.