Asymptotics and Extremal Properties of the Edge-Triangle Exponential Random Graph Model

Alessandro Rinaldo
Carnegie Mellon University
joint work with Mei Yin, Sukhada Fadnavis, Stephen E. Fienberg and Yi Zhou

May 20, 2014
Algebraic Statistics 2014
Illinois Institute of Technology
Outline

- Exponential random graphs models (ERGMs) for network data.
- Degeneracy and Geometry of Discrete exponential families.
- Asymptotic Geometry of the Edge-Triangle Model.
- Asymptotic Extremal Properties of the Edge Triangle Model.
Exponential random graphs models (ERGMs) for network data.

Degeneracy and Geometry of Discrete exponential families.

Asymptotic Geometry of the Edge-Triangle Model.

Asymptotic Extremal Properties of the Edge Triangle Model.

Asymptotic quantization of exponential random graphs (2103)
Yin, M., Rinaldo, A. and Fadnavis, S.
http://arxiv.org/abs/1311.1738
Let $\mathcal{G}_n$ be the set of simple graphs on $n$ nodes. Thus, $|\mathcal{G}_n| = 2^{\binom{n}{2}}$. 
Let \( \mathcal{G}_n \) be the set of simple graphs on \( n \) nodes. Thus, \( |\mathcal{G}_n| = 2^{\binom{n}{2}} \).

The nodes represent the units of some (sub)population of interest. The edges of any \( x \in \mathcal{G}_n \) encode a set of static relationships among population units.

Source: Chung and Lu (2006), Complex Graphs and Networks.
Let $\mathcal{G}_n$ be the set of simple graphs on $n$ nodes. Thus, $|\mathcal{G}_n| = 2^{(\binom{n}{2})}$.

The nodes represent the units of some (sub)population of interest. The edges of any $x \in \mathcal{G}_n$ encode a set of static relationships among population units.

Source: Chung and Lu (2006), Complex Graphs and Networks.
Exponential random graph models arise by specifying a set of informative network statistics on $G_n$

$$x \mapsto t(G_n) = (t_1(G_n), \ldots, t_d(G_n)) \in \mathbb{R}^d,$$

such that the probability of observing $G_n$ is a function of $t(G_n)$ only.
Exponential random graph models arise by specifying a set of informative network statistics on $G_n$

$$x \mapsto t(G_n) = (t_1(G_n), \ldots, t_d(G_n)) \in \mathbb{R}^d,$$

such that the probability of observing $G_n$ is a function of $t(G_n)$ only.

Examples:

- the number of edges $E(G_n)$ (Erdős-Renyi model);
- number of triangles $T(G_n)$;
- the number of k-stars;
- the degree sequence (the $\beta$-model) or degree distribution.
Exponential random graph models arise by specifying a set of informative network statistics on $G_n$

$$x \mapsto t(G_n) = (t_1(G_n), \ldots, t_d(G_n)) \in \mathbb{R}^d,$$

such that the probability of observing $G_n$ is a function of $t(G_n)$ only.

Examples:

- the number of edges $E(G_n)$ (Erdös-Renyi model);
- number of triangles $T(G_n)$;
- the number of k-stars;
- the degree sequence (the $\beta$-model) or degree distribution.
- ...or any combinations of statistics.
This talk: the Edge-Triangle (ET) Model

- **ET Model**: the network statistics are $t(G_n) = (E(G_n), T(G_n)) \in \mathbb{N}^2$. 

Example: \( n = 9 \). On \( G_9 \), there are 2,364 distinct graphs but only 444 distinct network statistics.
- **ET Model**: the network statistics are $t(G_n) = (E(G_n), T(G_n)) \in \mathbb{N}^2$.

- **Example**: $n = 9$. On $G_9$, there are $2^{36}$ distinct graphs but only 444 distinct network statistics.
Exponential Random Graph Models (ERGMs)

- **Exponential family representation** Given the choice $t$ of network statistics, construct the exponential family of probability distributions $\{Q_\theta, \theta \in \Theta\}$ on $G_n$ such that, for a given parameter $\theta \in \Theta$, the probability of observing $x$ is

$$Q_\theta(G_n) = \exp \{ \langle \theta, t(G_n) \rangle - \psi(\theta) \},$$
Exponential Random Graph Models (ERGMs)

- **Exponential family representation** Given the choice \( t \) of network statistics, construct the exponential family of probability distributions \( \{ Q_\theta, \theta \in \Theta \} \) on \( G_n \) such that, for a given parameter \( \theta \in \Theta \), the probability of observing \( x \) is

\[
Q_\theta(G_n) = \exp \left\{ \langle \theta, t(G_n) \rangle - \psi(\theta) \right\},
\]

where

- \( \psi(\theta) = \log \left( \sum_{x \in G_n} e^{\langle \theta, t(G_n) \rangle} \right) \) the log-partition function - a (often intractable) normalizing term;
- \( \Theta = \{ \theta \in \mathbb{R}^d : e^{\psi(\theta)} < \infty \} = \mathbb{R}^d \) is the natural parameter space.
Exponential Random Graph Models (ERGMs)

- **Exponential family representation** Given the choice $t$ of network statistics, construct the exponential family of probability distributions $\{Q_\theta, \theta \in \Theta\}$ on $G_n$ such that, for a given parameter $\theta \in \Theta$, the probability of observing $x$ is

$$Q_\theta(G_n) = \exp \{ \langle \theta, t(G_n) \rangle - \psi(\theta) \},$$

where

- $\psi(\theta) = \log \left( \sum_{x \in G_n} e^{\langle \theta, t(G_n) \rangle} \right)$ the log-partition function - a (often intractable) normalizing term;
- $\Theta = \{ \theta \in \mathbb{R}^d : e^{\psi(\theta)} < \infty \} = \mathbb{R}^d$ is the natural parameter space.

- **Two observations:**
  - It may be invariant with respect to relabeling of the vertices.
  - **Redundancy:** if $t(G'_n) = t(G_n)$, then $Q_\theta(G_n) = Q_\theta(G'_n)$, for all $\theta$. For the ET example, the median number of graphs corresponding to a network statistic is $2,741,130!$
Sufficiency Principle

Let $T_n = \{ t : t = t(G_n), G_n \in G_n \} \subset \mathbb{R}^d$ and
\[ \nu(t) = |\{ G_n \in G_n : t(G_n) = t \}|. \]

Consider instead the family of probability distributions $\{ P_\theta, \theta \in \Theta \}$ on $T_n$ such that, for a given parameter $\theta \in \Theta$, the probability of observing $t$ is
\[ P_\theta(t) = \exp \{ \langle \theta, t \rangle - \psi(\theta) \} \nu(t), \]
with the same $\Theta$ and $\psi(\theta)$. 
Sufficiency Principle

- Let $T_n = \{ t : t = t(G_n), G_n \in \mathcal{G}_n \} \subset \mathbb{R}^d$ and $\nu(t) = |\{ G_n \in \mathcal{G}_n : t(G_n) = t \}|$.

- Consider instead the family of probability distributions $\{ P_\theta, \theta \in \Theta \}$ on $T_n$ such that, for a given parameter $\theta \in \Theta$, the probability of observing $t$ is

$$P_\theta(t) = \exp \{ \langle \theta, t \rangle - \psi(\theta) \} \nu(t),$$

with the same $\Theta$ and $\psi(\theta)$.

- In the ET example, instead of $\mathcal{G}_9$ we can work with $T_n$.
Given one observation $x \in \mathcal{G}_n$, i.e. given $t = t(G_n)$,
  - estimate $\theta$ (usually with the MLE);
  - assess whether the ERG model fits the data.
Statistical Inference

- Given one observation \( x \in G_n, \) i.e. given \( t = t(G_n), \)
  - estimate \( \theta \) (usually with the MLE);
  - assess whether the ERG model fits the data.

- In the ET model, we want to learn 2 parameters: the edge and triangle parameters.
Despite being discrete exponential families, many ERGMs are troublesome and behave “extremely badly”.
ERGMs are Bad! ...I mean Hard!

- Despite being discrete exponential families, many ERGMs are troublesome and behave “extremely badly”.
- ERGMs and most network models are highly non-standard models for which traditional parametric statistics does not apply.
ERGMs are Bad! ...I mean Hard!

- Despite being discrete exponential families, many ERGMs are troublesome and behave “extremely badly”.
- ERGMs and most network models are highly non-standard models for which traditional parametric statistics does not apply.

The asymptotic (large $n$) properties of most ERGMs remain unknown.
ERGMs are Bad! ...I mean Hard!

- Despite being discrete exponential families, many ERGMs are troublesome and behave "extremely badly".
- ERGMs and most network models are highly non-standard models for which traditional parametric statistics does not apply.

The asymptotic (large $n$) properties of most ERGMs remain unknown.

- This talk will focus on extremal asymptotics of the ET model. See Chatterjee and Diaconis (2013) for many other asymptotic results.
- One route to the extremal asymptotics of the ET goes through degeneracy.
What is degeneracy?

- Quoting from Handcock (2003):
  - “A random graph model is near degenerate if the model places almost all its probability mass on a small number of graph configurations [...] e.g. empty graph, full graph, an individual graph, no 2-stars”;
  - a degenerate model is not “able to represent a range of realistic [networks]" since only a “small range of graphs [is] covered as the parameters vary”;

- the MLE does not exist and/or MCMCMLE fails to converge;
- the observed network \( t \) is very unlikely under the distribution specified by the MLE;
- the model misbehaves...
We capture overall degenerate behavior using Shannon’s entropy function

\[ \theta \mapsto - \sum_{t \in T_n} P_{\theta}(t) \log_2 \left( \frac{P_{\theta}(t)}{\nu(t)} \right), \quad \theta \in \Theta. \]

Rationale: degeneracy occurs when the probability mass is spread over a small number of network statistics, so degenerate distributions will tend to correspond to values of \( \theta \) for which the entropy function is small.

Entropy plot: for each point \( \theta \in \mathbb{R}^2 \) in the natural parameter space, plot the entropy of the corresponding probability distribution.
Degeneracy in the ET Model: Entropy Plots

Entropy plot: Natural Parameter Space

-10 -5 0 5 10 15 20 25
Edge parameter

-25 -15 -10 -5 0 5 10 15 20
Triangle parameter

0 5 10 15 20

Degeneracy and Geometry of Discrete Exponential Families
Asymptotic Geometry of the ET Model
Asymptotic/Extremal Properties of the ET Model
Degeneracy in the ET Model: Entropy Plots
Basics of Discrete Exponential Families

See Barndorff-Nielsen (1974) and Brown (1986)

- Let $T$ be a random vector taking values in a finite set $T_n$, for example $0, 5, 10, 15, 20, 25, 30, 35, 40$

  The distribution of $T$ belongs to the exponential family $\mathcal{E} = \{P_\theta, \theta \in \Theta\}$, with

  $$P_\theta(t) = \exp \{ \langle \theta, t \rangle - \psi(\theta) \} \nu(t), \quad \theta \in \Theta = \mathbb{R}^d.$$
Basics of Discrete Exponential Families

See Barndorff-Nielsen (1974) and Brown (1986)

- Let $T$ be a random vector taking values in a finite set $\mathcal{T}_n$, for example:

The distribution of $T$ belongs to the exponential family $\mathcal{E} = \{ P_\theta, \theta \in \Theta \}$, with

$$P_\theta(t) = \exp \{ \langle \theta, t \rangle - \psi(\theta) \} \nu(t), \quad \theta \in \Theta = \mathbb{R}^d.$$ 

- The set $\mathcal{P}_n = \text{convhull}(\mathcal{T}_n)$ is called the convex support.
Basics of Discrete Exponential Families

See Barndorff-Nielsen (1974) and Brown (1986)

- Let $T$ be a random vector taking values in a finite set $\mathcal{T}_n$, for example

The distribution of $T$ belongs to the exponential family $\mathcal{E} = \{P_\theta, \theta \in \Theta\}$, with

$$P_\theta(t) = \exp\{\langle \theta, t \rangle - \psi(\theta)\} \nu(t), \quad \theta \in \Theta = \mathbb{R}^d.$$ 

- The set $\mathcal{P}_n = \text{convhull}(\mathcal{T}_n)$ is called the convex support.

  - It is a polytope.

  - $\text{int}(\mathcal{P}_n) = \{E_\theta[\mathcal{T}], \theta \in \Theta\}$ is precisely the set of all possible expected values of $T$: mean value space.

  - $\text{int}(\mathcal{P}_n)$ and $\Theta$ are homeomorphic: we can represent the exponential family using $\text{int}(\mathcal{P}_n)$ instead of $\Theta$: mean value parametrization.
• Convex support for the ET example.
One-to-one correspondence between $\Theta$ and $\text{int}(\mathcal{P}_n)$. 
Extended Exponential Families: Geometric Construction

- For every face $F$ of $\mathcal{P}_n$, construct the exponential family of distributions $\mathcal{E}_F$ for the points in $F$ with convex support $F$. Note that $\mathcal{E}_F$ depends on $\dim(F) < d$ parameters only.
For every face $F$ of $\mathcal{P}_n$, construct the exponential family of distributions $\mathcal{E}_F$ for the points in $F$ with convex support $F$. Note that $\mathcal{E}_F$ depends on $\dim(F) < d$ parameters only.

The extended exponential family is

$$\overline{\mathcal{E}} = \mathcal{E} \cup \{\mathcal{E}_F : F \text{ is a face of } \mathcal{P}_n\}$$
For every face $F$ of $\mathcal{P}_n$, construct the exponential family of distributions $\mathcal{E}_F$ for the points in $F$ with convex support $F$. Note that $\mathcal{E}_F$ depends on $\dim(F) < d$ parameters only.

The extended exponential family is

$$\overline{\mathcal{E}} = \mathcal{E} \cup \{\mathcal{E}_F : F \text{ is a face of } \mathcal{P}_n\}$$

The extended exponential family is the closure of the original family. Geometrically, this corresponds to taking the closure of the mean value space, i.e. including the boundary of $\mathcal{P}_n$. 

- For every face $F$ of $\mathcal{P}_n$, construct the exponential family of distributions $\mathcal{E}_F$ for the points in $F$ with convex support $F$. Note that $\mathcal{E}_F$ depends on $\dim(F) < d$ parameters only.

- The extended exponential family is

$$\overline{\mathcal{E}} = \mathcal{E} \cup \{\mathcal{E}_F : F \text{ is a face of } \mathcal{P}_n\}$$

- The extended exponential family is the closure of the original family. Geometrically, this corresponds to taking the closure of the mean value space, i.e. including the boundary of $\mathcal{P}_n$. 

Graphical representation:

- Graph showing the extended exponential family with points and convex supports.
Back to Degeneracy

Entropy plot - Natural Parameter Space

Entropy plot - Mean Value Space

Entropy plot - Natural Parameter Space

Entropy plot - Mean Value Space
Let $\mathcal{P}$ be a full-dimensional polytope in $\mathbb{R}^d$. The normal cone to a face $F$ is the polyhedral cone

$$N_F = \left\{ c \in \mathbb{R}^k : F \subset \{ x \in \mathcal{P} : \langle c, x \rangle = \max_{y \in \mathcal{P}} \langle c, y \rangle, \} \right\}$$

consisting of all the linear functionals on $\mathcal{P}$ that are maximal over $F$. 
Let $\mathcal{P}$ be a full-dimensional polytope in $\mathbb{R}^d$. The normal cone to a face $F$ is the polyhedral cone

$$N_F = \left\{ c \in \mathbb{R}^k : F \subset \{ x \in \mathcal{P} : \langle c, x \rangle = \max_{y \in \mathcal{P}} \langle c, y \rangle \} \right\}$$

consisting of all the linear functionals on $\mathcal{P}$ that are maximal over $F$. The set of cones

$$\mathcal{N}(\mathcal{P}) = \{ N_F, F \text{ is a face of } \mathcal{P} \}$$

is called the normal fan of $\mathcal{P}$. 
Let $\mathcal{P}$ be a full-dimensional polytope in $\mathbb{R}^d$. The normal cone to a face $F$ is the polyhedral cone

$$N_F = \left\{ c \in \mathbb{R}^k : F \subset \{ x \in \mathcal{P} : \langle c, x \rangle = \max_{y \in \mathcal{P}} \langle c, y \rangle \} \right\}$$

consisting of all the linear functionals on $\mathcal{P}$ that are maximal over $F$. The set of cones

$$\mathcal{N}(\mathcal{P}) = \{ N_F, F \text{ is a face of } \mathcal{P} \}$$

is called the normal fan of $\mathcal{P}$.

Key properties:
- The (relative interiors of the) cones in $\mathcal{N}(\mathcal{P})$ partition $\mathbb{R}^d$.
- $\dim(N_F) = d - \dim(F)$. 
The Normal Fan for the ET Model
The Normal Fan for the ET Model

![Diagram of the Normal Fan](image)
The Normal Fan for the ET Model

Normal Fan

Number of edges

Number of triangles

−2 −1 0 1 2
−1
0
1
Normal Fan

0 20 40
−20
0
20
40
60
80
100
Number of edges

−20
0
20
40
60
80
100
Number of triangles

0 20 40
The Normal Fan for the ET Model

![Normal Fan Diagram]

- Number of edges
- Number of triangles

A. Rinaldo

Asymptotics and Extremal Properties of the ET Model
Degeneracy Explained (...Graphically)

- Entropy plots of the natural space and mean value spaces.
Entropy plots of the natural space space with superimposed the normal fan and of the mean value space.
Main Result (Colloquial Form)

- Pick any $\theta \in \Theta$, any face $F$ of $\mathcal{P}_n$ and any direction $o \neq 0$ in the interior of the normal cone $N_F$.
- For a sequence of positive numbers $r_i \to \infty$, form the sequence $\theta_i = \theta + r_i o$ in $\Theta$.
- Let $\mu_i = \mathbb{E}_{\theta_i}[T]$ be the mean value parameter corresponding to $\theta_i$; then, the sequence $\{\mu_i\}$ is contained in the interior of $\mathcal{P}_n$.
- Then, $\lim_{i} \mu_i$ is a point in the interior of $F$, which depends only on $\theta$ and $d$.
- Conversely, any point on the boundary of the convex support can be obtained in this way.
Main Result (Colloquial Form)

- Pick any $\theta \in \Theta$, any face $F$ of $P_n$ and any direction $o \neq 0$ in the interior of the normal cone $N_F$.
- For a sequence of positive numbers $r_i \to \infty$, form the sequence $\theta_i = \theta + r_i o$ in $\Theta$.
- Let $\mu_i = \mathbb{E}_{\theta_i}[T]$ be the mean value parameter corresponding to $\theta_i$; then, the sequence $\{\mu_i\}$ is contained in the interior of $P_n$.
- Then, $\lim_i \mu_i$ is a point in the interior of $F$, which depends only on $\theta$ and $d$.
- Conversely, any point on the boundary of the convex support can be obtained in this way.

Normal Fan and Extended Exponential Families

The normal fan realizes geometrically the closure of the family inside the natural parameter space.

See Rinaldo, Fienberg and Zhou (2009) and Geyer (2009)
Now let $n \to \infty$...
Now let $n \rightarrow \infty$...
Now let $n \to \infty$...
Now let $n \to \infty$...
Now let $n \to \infty$...
Now let $n \to \infty$. ...

We are interested in the following:

What happens when $\theta = (\theta_1, \theta_2)$ diverges to infinity along fixed directions and $n \to \infty$?
Now let $n \to \infty$...

We are interested in the following:

What happens when $\theta = (\theta_1, \theta_2)$ diverges to infinity along fixed directions and $n \to \infty$?

- This would correspond to taking the asymptotic closure of the ET model. Interpretation: how does the model behave when both $\|\theta\|$ and $n$ are large?

- We will require the use of fairly recent results on the theory of graph limits (though see Häggström and Jonasson, 1999).
For $G_n \in \mathcal{G}_n$ and a graph $H$ with $V(H) \leq n$ vertices, the density homomorphism of $H$ in $G_n$ is

$$t(H, G_n) = \frac{|\text{hom}(H, G_n)|}{|n^{|V(H)|}}. $$

In particular, $t(K_2, G_n) = \frac{2E(G_n)}{n^2}$ and $t(K_3, G_n) = \frac{6T(G_n)}{n^3}$. 
Rescaling First!

- For $G_n \in \mathcal{G}_n$ and a graph $H$ with $V(H) \leq n$ vertices, the density homomorphism of $H$ in $G_n$ is
  \[ t(H, G_n) = \frac{|\text{hom}(H, G_n)|}{|n| |V(H)|}. \]

- In particular, $t(K_2, G_n) = \frac{2E(G_n)}{n^2}$ and $t(K_3, G_n) = \frac{6T(G_n)}{n^3}$.

---

Rescaling of the ET model

To avoid trivialities, rescale the ET model so that

\[ Q_\theta(G_n) = \exp \left\{ n^2 \left( \langle \theta, t(G_n) \rangle - \psi_n(\theta) \right) \right\}, \quad G_n \in \mathcal{G}_n \]

where $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ and $t(G_n) = (t(K_2, G_n), t(K_3, G_n)) \in \mathbb{R}^2$.

This rescaling ensures that the sufficient statistics are of the same order $n^2$. 

A. Rinaldo  
Asymptotics and Extremal Properties of the ET Model  
21/36
For $G_n \in \mathcal{G}_n$ and a graph $H$ with $V(H) \leq n$ vertices, the density homomorphism of $H$ in $G_n$ is

$$t(H, G_n) = \frac{\|\text{hom}(H, G_n)\|}{n^{|V(H)|}}.$$ 

In particular, $t(K_2, G_n) = \frac{2E(G_n)}{n^2}$ and $t(K_3, G_n) = \frac{6T(G_n)}{n^3}$.

**Rescaling of the ET model**

Let $\mathcal{T}_n = \{t(G_n), G_n \in \mathcal{G}_n\} \subset [0, 1]^2$ and let $\mathcal{E}_n = \{P_{n, \theta}, \theta \in \mathbb{R}^2\}$ be the exponential family for $t(G_n)$:

$$P_{\theta}(t) = \exp \left\{ n^2(\langle \theta, t \rangle - \psi_n(\theta)) \right\} \nu_n(t), \quad t \in \mathcal{T}_n.$$ 

Also let $\mathcal{P}_n = \text{convhull}(\mathcal{T}_n) \subset [0, 1]^2$. 


Turán Graphs

- Turán graphs are fundamental in extremal graph theory and also in describing the asymptotic closure of the ET model.
Turán Graphs

- Turán graphs are fundamental in extremal graph theory and also in describing the asymptotic closure of the ET model.

- Let \( T(n, r) \) be a Turán graph on \( n \) nodes and \( r \) classes, \( r = 1, 2, \ldots, n \) (complete \( r \)-partite graph with partition sets differing in size by at most 1).

- For \( k = 0, 1, \ldots, n - 1 \), set

\[
\nu_{k,n} = t(T(n, k + 1)) = \begin{bmatrix}
t(K_2, T(n, k + 1)) \\
t(K_3, T(n, k + 1))
\end{bmatrix} \in [0, 1]^2.
\]
Turán Graphs

- Turán graphs are fundamental in extremal graph theory and also in describing the asymptotic closure of the ET model.

- Let $T(n, r)$ be a Turán graph on $n$ nodes and $r$ classes, $r = 1, 2, \ldots, n$ (complete $r$-partite graph with partition sets differing in size by at most 1).

- For $k = 0, 1, \ldots, n - 1$, set

  $$v_{k,n} = t(T(n, k + 1)) = \begin{bmatrix} t(K_2, T(n, k + 1)) \\ t(K_3, T(n, k + 1)) \end{bmatrix} \in [0, 1]^2.$$  

Vertices of $\mathcal{P}_n$ are Turán Graphs (Bollobás, 1976)

The vertices of $\mathcal{P}_n$ are the points $\{v_{k,n}, k = 0, 1, \ldots, \lceil n/2 \rceil - 1\}$ and $v_{n-1,n}$.  

Vertices of $\mathcal{P}_n$ are Turán Graphs (Bollobás, 1976)
For the asymptotics of the ET model we need the sets (well studied in extremal graph theory)

\[ T = \text{cl} \left( \{ t(G_n), G_n \in G_n, n = 1, 2, \ldots \} \right) \quad \text{and} \quad P = \text{convhull}(T). \]
For the asymptotics of the ET model we need the sets (well studied in extremal graph theory)

\[ T = \text{cl} \left( \{ t(G_n), G_n \in G_n, n = 1, 2, \ldots \} \right) \quad \text{and} \quad P = \text{convhull}(T). \]

**The set \( T \) (Razborov, 2008)**

\( T \) is the **closed** subset of \([0, 1]^2\) with

- upper boundary curve \( y = x^{3/2} \)
- and lower boundary given by \( y = 0 \) for \( 0 \leq x \leq 1/2 \) and

\[
y \geq \frac{(k - 1) \left( k - 2 \sqrt{k(k - x(k + 1))} \right) \left( k + \sqrt{k(k - x(k + 1))} \right)^2}{k^2(k + 1)^2}
\]

for \( (k - 1)/k \leq x \leq k/(k + 1), \ k = 2, 3, \ldots \).
Limiting Geometry of the ET Model
More on $\mathcal{T}$

- **Infinite homomorphism densities of Turán graphs:** For $k = 0, 1, \ldots$, let

$$
\lim_{n \to \infty} v_{k,n} = v_k := \left( \frac{k}{k+1}, \frac{k(k - 1)}{(k + 1)^2} \right) \in \mathbb{R}^2.
$$
More on $\mathcal{T}$

- **Infinite homomorphism densities of Turán graphs:** For $k = 0, 1, \ldots$, let
  \[
  \lim_{n \to \infty} v_{k,n} = v_k := \left( \frac{k}{k + 1}, \frac{k(k - 1)}{(k + 1)^2} \right) \in \mathbb{R}^2.
  \]

- Think of lower boundary of $\mathcal{T}$ as a series of “scallops” (strictly concave segments) between the $v_k$'s.
More on $\mathcal{T}$

- **Infinite homomorphism densities of Turán graphs:** For $k = 0, 1, \ldots$, let

  $$\lim_{n \to \infty} v_{k,n} = v_k := \left( \frac{k}{k+1}, \frac{k(k-1)}{(k+1)^2} \right) \in \mathbb{R}^2.$$ 

- Think of lower boundary of $\mathcal{T}$ as a series of "scallops" (strictly concave segments) between the $v_k$'s.
More on $\mathcal{T}$

triangle density $t$

$t = e^{3/2}$

$(0,0)$

$(1/2,0)$

edge density $e$

$t = e(2e - 1)$

$(1,1)$

E-R line
The set \( \mathcal{P} \) is defined as:

\[ \mathcal{P} = \lim_{n \to \infty} \mathcal{P}_n, \text{ with } \mathcal{P}_n \subset \mathcal{P} \text{ for all } n. \]

The extreme points of \( \mathcal{P} \) are \( \{v_k, k = 0, 1, \ldots\} \) and \( \lim_k v_k = (1, 1) \).

If \( n \) is a multiple of \( (k + 1)(k + 2) \), then \( v_{k,n} = v_k \) and \( v_{k+1,n} = v_{k+1} \). For all such \( n \), the line segment connecting \( v_k \) and \( v_{k+1} \) intersects \( \mathcal{P} \) only in \( v_k \) and \( v_{k+1} \).

Engström and Norén (2011) show a similar result using isomorphism densities, in which case \( \mathcal{P} = \bigcap_n \mathcal{P}_n \).
The set $\mathcal{P}$

The set $\mathcal{P} = \text{convhull}(\mathcal{T})$

- $\mathcal{P} = \lim_{n \to \infty} \mathcal{P}_n$, with $\mathcal{P}_n \subset \mathcal{P}$ for all $n$.
- The extreme points of $\mathcal{P}$ are $\{v_k, k = 0, 1, \ldots\}$ and $\lim_k v_k = (1, 1)$.
- If $n$ is a multiple of $(k + 1)(k + 2)$, then $v_{k,n} = v_k$ and $v_{k+1,n} = v_{k+1}$. For all such $n$, the line segment connecting $v_k$ and $v_{k+1}$ intersects $\mathcal{P}$ only in $v_k$ and $v_{k+1}$.

- Engström and Norén (2011) show a similar result using isomorphism densities, in which case $\mathcal{P} = \cap_n \mathcal{P}_n$. 
The set $\mathcal{P}$
The set $\mathcal{P}$
Let $L_k$ the line segment joining $v_k$ and $v_{k_1}$ and set $L_{-1}$ the segment joining $v_0 = (0,0)$ and $(1,1) = \lim_k v_k$. 

We call the $o_k$ the critical directions of the model.

Finally, set $C_k = \text{cone}(o_k - 1, o_k)$ for $k = 0, 1, 2, ...$. 

The Normal Fan to $\mathcal{P}$
Let $L_k$ the line segment joining $v_k$ and $v_{k_1}$ and set $L_{-1}$ the segment joining $v_0 = (0, 0)$ and $(1, 1) = \lim_k v_k$.

The outer normal to $L_k$ are

$$o_k = \begin{cases} 
(-1, 1) & \text{if } k = -1, \\
(0, -1) & \text{if } k = 0, \\
(1, -\frac{(k+1)(k+2)}{k(3k+5)}) & \text{if } k = 1, 2, \ldots
\end{cases}$$

We call the $o_k$ the critical directions of the model.
Let $L_k$ the line segment joining $v_k$ and $v_{k+1}$ and set $L_{-1}$ the segment joining $v_0 = (0, 0)$ and $(1, 1) = \lim_k v_k$.

The outer normal to $L_k$ are

$$o_k = \begin{cases} (-1, 1) & \text{if } k = -1, \\ (0, -1) & \text{if } k = 0, \\ \left(1, -\frac{(k+1)(k+2)}{k(3k+5)}\right) & \text{if } k = 1, 2, \ldots \end{cases}$$

We call the $o_k$ the critical directions of the model.

Finally, set $C_k = \text{cone}(o_{k-1}, o_k)$ for $k = 0, 1, \ldots$. 

The Normal Fan to $\mathcal{P}$
The Normal Fan to $P$
Recall that $P_{\theta,n}$ is the ET probability distribution on $\mathcal{T}_n$ corresponding to $\theta \in \mathbb{R}^2$. 
Asymptotics along generic directions

Pick arbitrary \( \theta, o \in \mathbb{R}^2 \) with \( o \neq 0 \) and \( o \notin \{ o_k, k = -1, 0, \ldots \} \) and let \( k \) be such that \( o \in C_k^o \). For any \( 0 < \epsilon < 1 \), there exists an \( n_0 = n_0(\theta, \epsilon, o) > 0 \) such that the following holds: for every \( n > n_0 \), there exists an \( r_0 = r_0(\theta, \epsilon, o, n) > 0 \) such that for all \( r > r_0 \),

\[
P_{\theta + r_0 n}(v_k, n) > 1 - \epsilon.
\]
Asymptotics along generic directions

Pick arbitrary $\theta, o \in \mathbb{R}^2$ with $o \neq 0$ and $o \notin \{ o_k, k = -1, 0, \ldots \}$ and let $k$ be such that $o \in C_k^o$. For any $0 < \epsilon < 1$, there exists an $n_0 = n_0(\theta, \epsilon, o) > 0$ such that the following holds: for every $n > n_0$, there exists an $r_0 = r_0(\theta, \epsilon, o, n) > 0$ such that for all $r > r_0$,

$$\mathbb{P}_{\theta + ro, n}(v_k, n) > 1 - \epsilon.$$ 

- **Choice of $\theta$ does not matter asymptotically.**
- **All points in $C_k^o$ yield the same asymptotics!**
- **When $n$ is a multiple of $(k + 1)(k + 2)$, this gives convergence in total variation (in $n$ and $r$, along subsequences) to $v_k$. For other $n$'s, we get weak convergence.**
Asymptotics along Generic Directions: Example

Directions that will give convergence to the full graph, empty graph and Turán graph with 2 classes
Technical assumption: \( n \) multiple of \((k + 1)(k + 2)\).
Most likely it can be removed.

Pick a \( o_k, k \neq \{-1, 0\} \). Let \( l_k \in \mathbb{R}^2 \) span the line through the origin parallel to \( L_k \) (line segment joining \( v_k \) and \( v_{k+1} \)) and consider the halfspaces

\[
H_k^+ = \{ x \in \mathbb{R}^2 : \langle x, l_k \rangle \geq 0 \} \quad \text{and} \quad H_k^- = \{ x \in \mathbb{R}^2 : \langle x, l_k \rangle < 0 \}.
\]
Asymptotics along critical directions

Let $\theta \in \mathbb{R}^2$, $k > 1$ and $0 < \epsilon < 1$, arbitrarily small. Then there exists an $n_0 = n_0(\theta, \epsilon, k) > 0$ such that the following holds: for every $n > n_0$ and a multiple of $(k + 1)(k + 2)$ there exists an $r_0 = r_0(\beta, \epsilon, k, n) > 0$ such that for all $r > r_0$,

- if $\theta \in H_k^+$ then $P_{\theta + r_0k, n}(v_{k+1}) > 1 - \epsilon$;
- if $\theta \in H_k^-$, then $P_{\theta + r_0k, n}(v_k) > 1 - \epsilon$. 
Asymptotics along critical directions

Let \( \theta \in \mathbb{R}^2 \), \( k > 1 \) and \( 0 < \epsilon < 1 \), arbitrarily small. Then there exists an \( n_0 = n_0(\theta, \epsilon, k) > 0 \) such that the following holds: for every \( n > n_0 \) and a multiple of \((k + 1)(k + 2)\) there exists an \( r_0 = r_0(\beta, \epsilon, k, n) > 0 \) such that for all \( r > r_0 \),

- if \( \theta \in H_k^+ \) then \( \mathbb{P}_{\theta + r_0, n}(v_{k+1}) > 1 - \epsilon \);
- if \( \theta \in H_k^- \), then \( \mathbb{P}_{\theta + r_0, n}(v_k) > 1 - \epsilon \).

- Asymptotics depends on whether \( \langle \theta, l_k \rangle \geq 0 \) or not.
- The model undergoes an asymptotic phase transition as \( \theta \) traverse the boundary of \( H_k^+ \). This type of discontinuity only occurs when \( n \to \infty \).
Asymptotics along Critical Directions. Example: $k = 2$. 

Graph showing edge density and triangle density.
The critical directions $o_{-1}$ and $o_0$ are different.
The Critical directions $o_{-1}$ and $o_0$

- The direction $o_{-1}$: For fixed $n$, the outer normal to the segment joining $v_{0,n}$ and $v_{n-1,n}$ is $o_{-1,n} := (-1, \frac{n}{n-2})$. For $n$ and $r$ large, the probability $P_{\theta + r o_{-1}, n}$, with $\theta = (\theta_1, \theta_2)$, assigns almost all of its mass to the empty graph when $\theta_1 + \theta_2 < 0$ and to the complete graph when $\theta_1 + \theta_2 > 0$, and it is uniform over $v_{0,n}$ and $v_{n-1,n}$ when $\theta_1 n(n-1) + \theta_2 (n-1)(n-2) = 0$.

- The direction $o_0$: For large $r$ and $n$, $P_{n, \theta + r o_0}$ with $\theta = (\theta_1, \theta_2)$, becomes indistinguishable from the exponential family on the number of edges of triangle-free graphs with natural parameter $\theta_1$. 
  More refined result: replace “triangle-free graphs" with “subgraphs of $T(n, 2)$." (See Diaconis and Chatterjee, 2013.)
We have shown that, when $n$ is a multiple of $(k + 2)(k + 1)$, $P_{n, \theta + r\theta}$ converges in total variation (weakly for other $n$'s) to point masses on the "corner points" and to a distribution on the horizontal segment of the set $T$, depending on $o$ and $\theta$, as $n, r \to \infty$ jointly.
We have shown that, when \( n \) is a multiple of \((k + 2)(k + 1)\), \( P_{n, \theta + r_o} \) converges in total variation (weakly for other \( n \)’s) to point masses on the “corner points” and to a distribution on the horizontal segment of the set 

\[
\mathcal{T} = \begin{align*}
\end{align*}
\]

depending on \( o \) and \( \theta \), as \( n, r \to \infty \) jointly.

Can this be extended to some form of stochastic convergence of the sequence \( \{G_{n,r}\} \) with \( G_{n,r} \sim P_{n, \theta + r_o} \)?
We have shown that, when \( n \) is a multiple of \((k + 2)(k + 1)\), \( P_{n,\theta + r\sigma} \) converges in total variation (weakly for other \( n \)’s) to point masses on the “corner points” and to a distribution on the horizontal segment of the set

\[
\mathcal{T} = \begin{array}{c}
0.0 & 0.2 & 0.4 & 0.6 & 0.8 & 1.0 \\
0.0 & 0.2 & 0.4 & 0.6 & 0.8 & 1.0 \\
\end{array}
\]

depending on \( \sigma \) and \( \theta \), as \( n, r \to \infty \) jointly.

Can this be extended to some form of stochastic convergence of the sequence \( \{G_{n,r}\} \) with \( G_{n,r} \sim P_{n,\theta + r\sigma} \)?

Yes, but we need graphons!
Let $\mathcal{W}$ be the space of measurable and symmetric functions from $[0, 1]^2$ into $[0, 1]$, called graphons. For $f \in \mathcal{W}$, the interval $[0, 1]$ can be thought as a continuum of vertices, and $f(x, y) = f(y, x)$ is the probability that $x$ and $y$ are connected by an edge.

Any finite graph $G_n$ has a graphon representation

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

The cut distance between graphons $f$ and $g$ is

$$d_\Box(f, g) = \sup_{S, T \subseteq [0, 1]} \left| \int_{S \times T} (f(x, y) - g(x, y)) \, dx \, dy \right|,$$

Let $\tilde{\mathcal{W}}$ arise from $\mathcal{W}$ by identifying graphons that are at a cut-distance 0 and related by a measure preserving transformation. The space $(\tilde{\mathcal{W}}, d_\Box)$ is compact.
A sequence \( \{ G_n \} \) of possibly random graphs converges to the graphon \( f \) when

\[
\lim_{n \to \infty} d_{\square}(f^{G_n}, f) = 0.
\]
Paraphrasing Our Results

There exists sequences $\{n_i, r_i\}_i$ such that the sequence of random graphs $\{G_i\}_i$ with $G_i \sim P_{n_i, \theta + r_i o}$ satisfies

$$\lim_i d_{\square}(f^{G_i}, f^{K_r}) = 0$$

in probability, where $f^{T_r}$ is the Turán graphon on $r$ classes

$$f^{T_r}(x, y) = \begin{cases} 
1 & \text{if } [xr] \neq [yr], \\
0 & \text{otherwise,}
\end{cases} \quad (x, y) \in [0, 1]^2,$$

with $r$ depending on $o$ and possibly $\theta$ ($[x]$ denotes the integer part of $x$).

When $o = o_0$, the limiting graphon is $p f^{T_2}$ with $p \in (0, 1)$ depending on $\theta_1$. 
Most of our results can also be derived (in a slightly different form) using a powerful analytic method due to Chatterjee and Diaconis (2013). However, it gives less precise results along critical directions.
Most of our results can also be derived (in a slightly different form) using a powerful analytic method due to Chatterjee and Diaconis (2013). However, it gives less precise results along critical directions.

Chatterjee and Diaconis (2013) showed that, for fixed, $\theta = (\theta_1, \theta_2)$, the limiting behavior of $\{G_n\}$ with $G_n \sim Q_{n, \theta}$, is determined by the solution of the variational problem

$$\sup_{f \in \tilde{W}} \left( \theta_1 t(K_2, f) + \theta_2 t(K_3, f) - \int \int_{[0,1]^2} I(f) dx dy \right),$$

where $I(u) = \frac{1}{2} u \log u + \frac{1}{2} (1 - u) \log(1 - u)$, and, for any finite graph $H$ on $k$ node,

$$t(H, f) := \int_{[0,1]^k} \prod_{\{i,j\} \in E(H)} f(x_i, x_j) dx_1 \cdots x_k.$$
Most of our results can also be derived (in a slightly different form) using a powerful analytic method due to Chatterjee and Diaconis (2013). However, it gives less precise results along critical directions.

Chatterjee and Diaconis (2013) showed that, for fixed, \( \theta = (\theta_1, \theta_2) \), the limiting behavior of \( \{G_n\} \) with \( G_n \sim Q_{n,\theta} \), is determined by the solution of the variational problem

\[
\sup_{f \in \tilde{W}} \left( \theta_1 t(K_2, f) + \theta_2 t(K_3, f) - \int \int_{[0,1]^2} l(f) dx dy \right),
\]

where \( l(u) = \frac{1}{2} u \log u + \frac{1}{2} (1 - u) \log(1 - u) \), and, for any finite graph \( H \) on \( k \) node,

\[
t(H, f) := \int_{[0,1]^k} \prod_{\{i,j\} \in E(H)} f(x_i, x_j) dx_1 \cdots x_k.
\]

This result can be applied to the case \( ||\theta|| \to \infty \) along fixed directions.
Conclusions

- We have fully described the asymptotic extremal behavior of the ET model.

- We made heavy use of recent results on graph limits (see, e.g., the book by Lovász, 2012 for details) and probability (Chatterjee and Diaconis, 2013 and Chatterjee and Varadhan, 2011).

- Some of our results and techniques are likely to generalize to more complex ERGMs.