Random graphs with a given degree sequence

Sourav Chatterjee  
(NYU)

Persi Diaconis  
(Stanford)

Allan Sly  
(Microsoft)
Let $G$ be an undirected simple graph on $n$ vertices.

Let $d_1, \ldots, d_n$ be the degrees of the vertices of $G$ arranged in descending order.

The vector $\mathbf{d} := (d_1, \ldots, d_n)$ is called the degree sequence of $G$. Equivalently, one can consider the degree distribution, i.e. the probability measure that puts mass $1/n$ at each $d_i$.

Vast interest in degree distributions of real-world graphs in recent times.

One approach: study graphs that are chosen uniformly from the set of all graphs on a given set of vertices with a given degree sequence.

What does such a random graph ‘look like’?

We give a rather precise answer to this question in the dense case, i.e. where the degrees are comparable to the number of vertices.
Some remarks

- Real world graphs are usually sparse; this is a gap between our theory and the practical situation.
- There is a closely related line of work due to Barvinok & Hartigan, with important similarities and differences (will come back to this later).
- There is intensive use of statistical methodology in our work, which is otherwise graph-theoretic in nature. Hence suitable for a statistical audience.
Towards a precise formulation of the question

- Let $F_n$ be a degree distribution for an $n$-vertex graph, normalized by $n$ so that it is supported in $[0, 1]$.
- Let $G_n$ be a (uniformly chosen) random graph with degree distribution $F_n$.
- Suppose that $F_n$ converges to a limit distribution $F$ as $n$ tends to infinity.
- What is the ‘limit’ of $G_n$?
- To answer this question, we need a theory of graph limits.
An abstract topological space of graphs


- Let $G_n$ be a sequence of simple graphs whose number of nodes tends to infinity.

- For every fixed simple graph $H$, let $|\text{hom}(H, G)|$ denote the number of homomorphisms of $H$ into $G$ (i.e. edge-preserving maps $V(H) \to V(G)$, where $V(H)$ and $V(G)$ are the vertex sets).

- This number is normalized to get the homomorphism density $t(H, G) := \frac{|\text{hom}(H, G)|}{|V(G)| |V(H)|}$.

This gives the probability that a random mapping $V(H) \to V(G)$ is a homomorphism.
Suppose that \( t(H, G_n) \) tends to a limit \( t(H) \) for every \( H \).

Then Lovász & Szegedy proved that there is a natural “limit object” in the form of a function \( f \in \mathcal{W} \), where \( \mathcal{W} \) is the space of all measurable functions from \([0, 1]^2\) into \([0, 1]\) that satisfy \( f(x, y) = f(y, x) \) for all \( x, y \).

Conversely, every such function arises as the limit of an appropriate graph sequence.

This limit object determines all the limits of subgraph densities: if \( H \) is a simple graph with \( k \) vertices, then

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

A sequence of graphs \( \{G_n\}_{n \geq 1} \) is said to converge to \( f \) if for every finite simple graph \( H \),

\[
\lim_{n \to \infty} t(H, G_n) = t(H, f).
\]
Example

For any fixed graph $H$,\[ t(H, G(n, p)) \to p|E(H)| \quad \text{almost surely as } n \to \infty. \]

On the other hand, if $f$ is the function that is identically equal to $p$, then $t(H, f) = p|E(H)|$.

Thus, the sequence of random graphs $G(n, p)$ converges almost surely to the non-random limit function $f(x, y) \equiv p$ as $n \to \infty$. 
The elements of $\mathcal{W}$ are sometimes called ‘graphons’.

A finite simple graph $G$ on $n$ vertices can also be represented as a graphon $f^G$ is a natural way:

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

Note that this allows all simple graphs, irrespective of the number of vertices, to be represented as elements of the single abstract space $\mathcal{W}$.

So, what is the topology on this space?
The cut metric

For any \( f, g \in \mathcal{W} \), Frieze and Kannan defined the cut distance:

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

Introduce an equivalence relation on \( \mathcal{W} \): say that \( f \sim g \) if \( f(x, y) = g_\sigma(x, y) := g(\sigma x, \sigma y) \) for some measure preserving bijection \( \sigma \) of \( [0, 1] \).

Denote by \( \tilde{g} \) the closure in \( (\mathcal{W}, d_{\square}) \) of the orbit \( \{g_\sigma\} \).

The quotient space is denoted by \( \tilde{\mathcal{W}} \) and \( \tau \) denotes the natural map \( g \rightarrow \tilde{g} \).

Since \( d_{\square} \) is invariant under \( \sigma \) one can define on \( \tilde{\mathcal{W}} \) the natural distance \( \delta_{\square} \) by

\[
\delta_{\square}(\tilde{f}, \tilde{g}) := \inf_{\sigma} d_{\square}(f, g_\sigma) = \inf_{\sigma} d_{\square}(f_\sigma, g) = \inf_{\sigma_1, \sigma_2} d_{\square}(f_{\sigma_1}, g_{\sigma_2})
\]

making \( (\tilde{\mathcal{W}}, \delta_{\square}) \) into a metric space.
To any finite graph $G$, we associate the natural graphon $f^G$ and its orbit $\tilde{G} = \tau f^G = \tilde{f}^G \in \tilde{\mathcal{W}}$. One of the key results of the is the following:

**Theorem (Borgs, Chayes, Lovász, Sós & Vesztergombi)**

A sequence of graphs $\{G_n\}_{n \geq 1}$ converges to a limit $f \in \mathcal{W}$ if and only if $\delta(\tilde{G}_n, \tilde{f}) \to 0$ as $n \to \infty$.

**Remark:** Besides subgraph counts, many other interesting functions are continuous with respect to this topology, e.g. see the survey by Austin & Tao.
Scaling limit of degree sequences

- Suppose that for each $n$, a degree sequence $d^n = (d^n_1, \ldots, d^n_n)$ is given, where $d^n_1 \geq d^n_2 \geq \cdots \geq d^n_n$.
- Suppose that there is a non-increasing function $f$ on $[0,1]$ such that

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left| \frac{d^n_i}{n} - f\left(\frac{i}{n}\right) \right| = 0,
$$

and $d^n_1/n \to f(0), \ d^n_n/n \to f(1)$.
- The first condition is equivalent to: $D_n/n \to f(U)$ in distribution, where $D_n$ is a randomly (uniformly) chosen $d^n_i$ and $U$ is uniformly distributed on $[0,1]$.
- The need to control $d^n_1$ and $d^n_n$ arises from the need to eliminate ‘outlier’ vertices that connect to almost all nodes or almost no nodes.
The space of scaling limits of degree sequences

Let \( D'[0,1] \) denote the set of all left-continuous non-increasing functions on \([0,1]\), endowed with the topology induced by a modified \( L^1 \) norm:

\[
\|f\|_{1'} := |f(0)| + |f(1)| + \int_0^1 |f(x)| \, dx.
\]

The set of all possible scaling limits of degree sequences is a subset of \( D'[0,1] \). Let \( \mathcal{F} \) denote this set.

Then \( \mathcal{F} \) is a closed subset of \( D'[0,1] \) under the modified \( L^1 \) norm.

Moreover, \( \mathcal{F} \) has non-empty interior, and the interior can be described in an explicit form.
The interior of $F$

From previous slide: $F$ is the space of all possible scaling limits of degree sequences, carrying the topology of the modified $L^1$ norm.

Proposition (Chatterjee-Diaconis-Sly)

A function $f : [0, 1] \rightarrow [0, 1]$ in $D'[0, 1]$ belongs to the interior of $F$ if and only if

(i) there are two constants $c_1 > 0$ and $c_2 < 1$ such that $c_1 \leq f(x) \leq c_2$ for all $x \in [0, 1]$, and

(ii) for each $x \in (0, 1]$,

$$\int_x^1 \min\{f(y), x\} \, dy + x^2 - \int_0^x f(y) \, dy > 0.$$
Suppose $d_1 \geq d_2 \geq \cdots \geq d_n$ are nonnegative integers.

The Erdős-Gallai criterion says that $d_1, \ldots, d_n$ is the degree sequence of a simple graph on $n$ vertices if and only if

$$\sum_{i=1}^n d_i$$

is even and for each $1 \leq k \leq n$,

$$k(k - 1) + \sum_{i=k+1}^n \min\{d_i, k\} - \sum_{i=1}^k d_i \geq 0.$$

A degree sequence is in the interior of the convex hull of all possible length-$n$ degree sequences if and only if strict inequality holds for all $k$.

The Proposition in the previous slide is a limiting version of the Erdős-Gallai criterion.
The main theorem

Theorem (Chatterjee-Diaconis-Sly)

Let $G_n$ be a random graph with given degree sequence $d^n$. Suppose $d^n$ converges to a scaling limit $f$ and suppose that $f$ belongs to the interior of $\mathcal{F}$. Then there exists a unique function $g : [0, 1] \rightarrow \mathbb{R}$ in $D'[0, 1]$ such that the function

$$W(x, y) := \frac{e^{g(x)+g(y)}}{1 + e^{g(x)+g(y)}},$$

satisfies, for all $x \in [0, 1]$,

$$f(x) = \int_{0}^{1} W(x, y) dy.$$

The sequence $\{G_n\}$ converges almost surely to the limit graph represented by the function $W$. 
Given $\beta = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n$, let $P_\beta$ be the law of the undirected random graph on $n$ vertices defined as follows: for each $1 \leq i \neq j \leq n$, put an edge between the vertices $i$ and $j$ with probability

$$p_{ij} := \frac{e^{\beta_i + \beta_j}}{1 + e^{\beta_i + \beta_j}},$$

independently of all other edges. We call this the $\beta$-model.

Then if $G$ is a graph with degree sequence $d_1, \ldots, d_n$, the probability of observing $G$ under $P_\beta$ is

$$e^{\sum_i \beta_i d_i} / \prod_{i<j} (1 + e^{\beta_i + \beta_j}).$$

This model was considered by Holland & Lienhardt in the directed case, by Park & Newman and Blitzstein & Diaconis in the undirected case. Similar to the Bradley-Terry model for rankings. See Hunter (2004) for extensive references.

It is also a simple version of a host of exponential models actively in use for analyzing network data.
Suppose $G$ is a random graph with given a degree sequence $d$.

**Step 1:** If we can find a well-behaved $\beta$ such that the $d$ is the expected degree sequence under the $\beta$-model, then a random graph drawn from the $\beta$-model is close to $G$ in the cut metric with high probability.

**Step 2:** If $d$ is away from the boundary of the Erdős-Gallai polytope, then there exists such a $\beta$.

**Step 3:** The $\beta$-model approximates the graph limit described in our main theorem if $n$ is large.

The proofs of all three steps are quite involved.

In a sequence of papers produced at the same time as our work, Barvinok & Hartigan established Steps 1 and 3, although in a different language. We work out all three steps.

On the other hand, the Barvinok-Hartigan results give finite sample error bounds and very precise asymptotics, while we only have limit theorems.
Maximum likelihood estimation in the $\beta$-model

Suppose a random graph $G$ is generated from the $\beta$-model, where $\beta \in \mathbb{R}^n$ is unknown.

The ML equations for $\beta$ are:

$$d_i = \sum_{j \neq i} \frac{e^{\hat{\beta}_i + \hat{\beta}_j}}{1 + e^{\hat{\beta}_i + \hat{\beta}_j}}, \quad i = 1, \ldots, n,$$

where $d_1, \ldots, d_n$ are the degrees in the observed graph $G$.

**Theorem (Chatterjee-Diaconis-Sly)**

Let $L := \max_{1 \leq i \leq n} |\beta_i|$. There is a constant $C(L)$ depending only on $L$ such that with probability at least $1 - C(L)n^{-2}$, there exists a unique solution $\hat{\beta}$ of the ML equations, that satisfies

$$\max_{1 \leq i \leq n} |\hat{\beta}_i - \beta_i| \leq C(L)\sqrt{n^{-1}\log n}.$$

Thus, all $n$ parameters may be estimated from a sample of size one!
Some remarks

Similar consistency result for the Bradley-Terry model due to Simons & Yao (1999).
Possible to get such a result because there are $n(n - 1)/2$ independent random variables lurking in the background.
Question: Is it possible to solve the ML equations quickly and deterministically?
Answer: Yes, we have an easy algorithm.
The algorithm is actually an important component of the proofs of the graph limit theorem and the consistency theorem.
The algorithm

- Let $d_1, \ldots, d_n$ be the degrees in a particular realization of a random graph from a $\beta$-model.

- For $1 \leq i \leq n$ and $x \in \mathbb{R}^n$, let
  \[
  \varphi_i(x) := \log d_i - \log \sum_{j \neq i} \frac{1}{e^{-x_j} + e^{x_i}}.
  \]

- Let $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ be the function whose $i$th component is $\varphi_i$.

Theorem (Chatterjee-Diaconis-Sly)

Suppose the ML equations have a solution $\hat{\beta}$. Then $\hat{\beta}$ is a fixed point of $\varphi$ and may be reached geometrically fast from any point in $\mathbb{R}^n$ by iterative applications of $\varphi$. If the ML equations do not have a solution, then the iterates must have a divergent subsequence.
Plot of $\hat{\beta}_i$ versus $\beta_i$ for a graph with 100 vertices, where $\beta_1, \ldots, \beta_n$ were chosen i.i.d. $\sim Unif[-1, 1]$. 
Simulations with a larger graph

Plot of $\hat{\beta}_i$ versus $\beta_i$ for a graph with 300 vertices, where $\beta_1, \ldots, \beta_n$ were chosen i.i.d. $\sim Unif[-1,1]$. The increased accuracy for larger $n$ is clearly visible.