Models of Real-world Networks: Inhomogeneous Random Graphs and Convergent Graph Sequences

Béla Bollobás
University of Cambridge and University of Memphis

b.bollobas@dpmms.cam.ac.uk

PLAN

I. The Classical Models
of Random Graphs

The theory of random graphs was founded
by **Erdős and Rényi** in the late 1950s and early 1960s.

$G_{n,m}$**:** random graph on $[n] = \{1, \ldots, n\}$, with m
edges selected at random (studied by ER);

$G_{n,p}$**:** random graph on $[n] = \{1, \ldots, n\}$, with edges
selected with probability p, independently of each other
(introduced by Gilbert).

For $p = m/\binom{n}{2}$ the two models are just about equivalent.

Giant Component – Percolation

$G_{n,p}$:   'mean field model'

percolation $\leftrightarrow$ giant component

(open cluster with cn sites)

critical probability

$\leftrightarrow$ threshold for a giant component

Erdős and Rényi 1961: $p_c = (1 + o(1))/n$.

Further Classical Models

$\mathcal{G}(n, r - \mathrm{reg})$:
the space or r-regular graphs (every degree is r)

and $\mathcal{G}(n, k - \mathrm{out})$:
first every vertex 'sends out' k edges chosen at random, the graph is formed by all these edges.

All these classical random graphs are **homogeneous**: their vertices are 'interchangeable'; for p 'small', $p \sim c/n$, say, the degree sequence has close to Poisson distribution.

II. Inhomogeneous Random Graphs

Watts and Strogatz (1998):
Large-scale real-world graphs look 'random' (no clearly discernible structure), but are **not** similar to the Erdős–Rényi graphs (or any of the widely studied random graphs): **they are far from homogeneous**.

E.g., in many real-life graphs, the **diameter** (maximal distance between vertices) **is small**
(as in a 'classical' random graph),

but the **'clustering coefficient'** (the probability that your friend's friend is your friend) is **much larger** than in a classical random graph.

Some Large-Scale Real-life Networks

- The Graph of the World-Wide Web

- The Internet Graph

- Networks of Scientific Citations

- Scientific Collaboration Networks

- The Network of Film Actors

- The Word Web of a Language

- Food Webs

- Numerous Biological Networks

Power Law

Barabási and Albert and
Faloutsos, Faloutsos and Faloutsos:

'power law in networks' (e.g., of the 'internet topology')

Related models:
Barabási, Albert and Jeong; Kumar, Raghavan, Rajagopalan, Sivakumar, Tomkins and Upfal; Broder, Kumar, Stata and many others

Scale-free nature is emphasized – power law for degree distribution

Networks and Models

In the last decade: **much work on large-scale real-world networks, and their models.**

- Direct studies of the real-world networks themselves, measuring various properties such as degree-distribution, diameter, clustering, etc.

- Constructions of new random graph models motivated by this study.

- Computer experiments on the new models, measuring their properties.

- Heuristic analysis of the new models to predict their properties.

- **Rigorous mathematical study of these models, to prove theorems about them.**

Aim

Construct models of inhomogeneous random graphs which are flexible and general, and are accessible to rigorous mathematical analysis.

Main questions: degree sequence, diameter, clustering coefficient, robustness (resistance to random breakdown), vulnerability (resistance to a malicious attack), etc.

[A system is up an running if there is a giant component.]

A model is 'satisfactory' if it gives the 'right answers to questions like these.

III. Two Examples

III.1. The Model of Dubins

In 1984 Dubins introduced the random graph $G_n(c)$ for $0 < c < \infty$ and $n \geq 1$.

Vertex set: $[n] = \{1, \ldots, n\}$.

Each pair ij is an edge independently of all other events, with probability c/j (really, $\min\{\mathbf{c/j}, \mathbf{1}\}$).

Question: for what values of c does $G_n(c)$ have a giant component?

III.2. The CHKNS model

In 1991 Callaway, Hopcroft, Kleinberg, Newman and Strogatz introduced (essentially) the foll. model.

At each time step, a new vertex is added.

At time t (when we have t vertices), join each pair of vertices with probability $\delta / \binom{t}{2}$.

The edges chosen are independent of each other for different pairs and/or different times.

$F_n(\delta) =$ the graph obtained at time **n.**

Also, $F_n(\delta)$ and $G_n(2\delta)$ are similar.

Results

Kalikow and Weiss 1988: whp (with high prob.) for $c < 1/4$ the graph $G_n(c)$ has only small components, while for $c > 1$ it has a giant component.

The conjecture remained: the critical value is 1.

Results

Kalikow and Weiss 1988: whp (with high prob.) for $c < 1/4$ the graph $G_n(c)$ has only small components, while for $c > 1$ it has a giant component.

The conjecture remained: the critical value is 1.

Surprise: in 1989, Shepp proved that for $c > 1/4$ the graph $G_n(c)$ has a giant component.

Thus for G_c the critical constant is 1/4.

A similar result for a more general model was proved by Durrett and Kesten in 1990.

Sharp Results

A number of precise results have been **proved** by Durrett, and B., Janson and Riordan.

Oliver Riordan has proved the following difficult and surprising result:

As $\varepsilon \longrightarrow 0$, whp we have

$$C_1(G_n(1/4 + \varepsilon)) = \exp\left(-\frac{1 + o(1)}{2\sqrt{\varepsilon}}\right) n.$$

Infinite order phase transition with rather precise speed.

Simple Inhomogeneous Models

Inhomogeneous random graphs on $V = [n]$:

1. Finite type graphs: $V = V_1 \cup V_2 \cup V_3$,

$\mathbb{P}(i \sim j) = \frac{c_{hk}}{n}$**, for** $i \in V_h$ **and** $j \in V_k$**.**

2. LCD model: rather close to $\mathbb{P}(i \sim j) = \frac{1}{\sqrt{ij}}$**.**

3. The Dubins and CHKNS models:

$\mathbb{P}(i \sim j) = \frac{1}{j}$ **if** $i < j$**.**

When is there a giant component?
Robustness? Vulnerability?

IV. THE BJR MODEL
of B, Janson and Riordan

This is a very general model of random graphs defined in 2007 that includes exactly many of the specific spaces of inhomogeneous random graphs that have been studied.

BJR proved precise results about the component structure of a random graph in this model by relating it to the survival probability of a multitype branching process, and the norm of a certain operator.

We also proved detailed results about the phase transition in this model, especially the numbers of vertices and edges in the giant component, and

the stability of the giant component under the addition and deletion of edges.

IV.1. The Construction

Let $\mathcal{S}$ be a separable metric space equipped with a Borel probability measure μ.
(Main example: $\mathcal{S} = (0,1]$ and μ is the Lebesgue measure.)

Kernel on $\mathcal{S}$: symmetric non-negative function on $\mathcal{S} \times \mathcal{S}$.

For each n we have a deterministic or random sequence $\mathbf{x}_n = (x_1, \ldots, x_{N_n})$ of points in $\mathcal{S}$ such that the empirical distribution

$$\mu_n = \frac{1}{n} \sum_{i=1}^{N_n} \delta_{x_i}$$

converges in probability to μ as $n \to \infty$, with convergence in the usual space of probability measures on $\mathcal{S}$.

We define a random graph $G_n(\kappa) = G^{\mathcal{V}}(n, \kappa)$ on $\{1, 2, \ldots, N_n\}$ by taking the probability that ij is an edge to be $\kappa(x_i, x_j)/n$, with all edges chosen independently.

Think of $G_n(\kappa)$ as a sparse inhomogeneous random graph with some edge-probabilities much larger than their average.

Our main interest is in the structure (especially the component structure) of the obtained random graph for large values of n.

Take a random graph, and then consider percolation on it.

IV.2. Branching Processes

To study the components of $G(n, \kappa)$, we use the multi-type Galton–Watson branching process with type space $\mathcal{S}$, where a particle of type $x \in \mathcal{S}$ is replaced in the next generation by a set of particles distributed as a Poisson process on $\mathcal{S}$ with intensity $\kappa(x, y) \, d\mu(y)$.

(Thus, the number of children with types in a subset $A \subseteq \mathcal{S}$ has a Poisson distribution with mean $\int_A \kappa(x, y) \, d\mu(y)$).

Write $\mathfrak{X}_\kappa$ for this branching process, started with a single particle chosen at random, distributed according to μ.

$\rho(\kappa)$ is the probability that the branching process $\mathfrak{X}_\kappa$ survives for eternity.

Operators

Let T_κ be the integral operator on $(\mathcal{S}, \mu)$ with kernel κ, defined by
$$(T_\kappa f)(x) = \int_{\mathcal{S}} \kappa(x, y) f(y) \, d\mu(y)$$

for any non-negative (measurable) function f.
[Thus this integral is defined (finite or $+\infty$) for a.e. x.]

The norm:

$$\|T_\kappa\| = \sup \left\{ \|T_\kappa f\|_2 : f \geq 0, \|f\|_2 \leq 1 \right\} \leq \infty.$$

IV.3. Results

Theorem. **(The Giant Component:**

existence, size, uniqueness)

Let (κ_n) **be a graphical sequence of kernels on**

a vertex space $\mathcal{V}$ **with limit** κ**.**

If $\|T_\kappa\| \leq 1,$ **then** $C_1\big(G^{\mathcal{V}}(n, \kappa_n)\big) = o_p(n)$**;**

if $\|T_\kappa\| > 1,$ **then** $C_1\big(G^{\mathcal{V}}(n, \kappa_n)\big) = \Theta(n)$ **whp.**

Thm ctd. For any $\varepsilon > 0$, whp we have
$$\frac{1}{n} C_1\big(G^{\mathcal{V}}(n, \kappa_n)\big) \le \rho(\kappa) + \varepsilon.$$

If κ is irreducible, then
$$\frac{1}{n} C_1(G^{\mathcal{V}}(n, \kappa_n)) \xrightarrow{\mathrm{p}} \rho(\kappa),$$

and
$$C_2(G^{\mathcal{V}}(n, \kappa_n)) = o(n).$$

In all cases $\rho(\kappa) < 1$; furthermore, $\rho(\kappa) > 0$ if and only if $\|T_\kappa\| > 1$.

Corollary. Let κ be a graphical kernel on a vertex space $\mathcal{V}$, and define the random graph $G = G^{\mathcal{V}}(n, \kappa)$. Consider bond percolation on G. Then the critical probability p_c is $\|T_\kappa\|^{-1}$.

More precisely, if $p \leq \|T_\kappa\|^{-1}$, then

$$C_1\big(G^{\mathcal{V}}(n, p\kappa)\big) = o_p(n),$$

while if $p > \|T_\kappa\|^{-1}$ and κ is irreducible, then

$$C_1\big(G^{\mathcal{V}}(n, p\kappa)\big) = \rho(p\kappa)n + o_p(n) = \Theta_p(n).$$

Corollary. **Let** κ **be a graphical kernel on a vertex space** $\mathcal{V}$. **Then**

$$\forall\, c > 0, \ \ G^{\mathcal{V}}(n, c\kappa) \text{ has a giant comp. whp}$$

iff

$$\|T_\kappa\| = \infty.$$

Otherwise the existence of a giant component has a finite threshold $c_0 > 0$.

QUESTIONS REMAIN

How can we tell that a space of random graphs models a certain large-scale real-world graph?

Degree sequence? Diameter? Robustness?

Other properties?

Not enough!

We need a METRIC on the space of finite graphs which ties our given graphs to the models.

V. DENSE GRAPHS

Borgs, Chayes, Lovász, Sós, Szegedy and Vesztergombi – several substantial papers

Given graphs F and G, $X_F(G)$ is the number of (not nec. ind.) copies of F in G.

The number of embeddings of F into G is
$$\mathrm{emb}(F, G) = \mathrm{aut}(F) X_F(G).$$

With $k = |F|$ and $n = |G|$,
$$s(F, G) = \frac{\mathrm{emb}(F,G)}{n_{(k)}} = \frac{X_F(G)}{X_F(K_n)} \in [0, 1].$$
is the normalized F-count.

Subgraph Distance

$\mathcal{F} = \{F_1, F_2, \ldots\}$: **isom. classes of fin. grs.**

s **maps** $\mathcal{F}$ **into** $[0,1]^\infty$ **(or into** $[0,1]^{\mathcal{F}}$**):**

$$s(G) = (s_i(G))_{i=1}^\infty \in [0,1]^\infty,$$

where $s_i(G) = s(F_i, G)$.

d: **any metric on** $[0,1]^\infty$ **inducing the product topology; e.g.,** $d(s,t) = \sum 2^{-i}|s_i - t_i|$.

The subgraph distance of two graphs G_1, G_2 **is**

$$d_{\mathrm{sub}}(G_1, G_2) = d(s(G_1), s(G_2)).$$

Kernels – Graphons

Kernel (graphon): a symmetric measurable function from $[0,1]^2$ to $[0,1]$.

graph $G \longrightarrow$ **kernel** κ_G

Given a finite graph F **with vertex set** $\{1, 2, \ldots, k\}$, **and a kernel** κ,

$$s(F, \kappa) = \int_{[0,1]^k} \prod_{ij \in E(F)} \kappa(x_i, x_j) \prod_{i=1}^{k} \mathrm{d}x_i.$$

Gives an extension of d_{sub} **to kernels.**

Cauchy Sequences

Theorem (Lovász and Szegedy, 2006)

Let (G_n) be a Cauchy sequence in $(\mathcal{F}, d_{\text{sub}})$ with $|G_n| \to \infty$. Then there is a kernel κ such that $s(F, G_n) \to s(F, \kappa)$.

The completion of $(\mathcal{F}, d_{\text{sub}})$ is obtained by adding to $\mathcal{F}$ the set $\mathcal{K}$ of all equivalence classes of kernels, and using the map

$s : \mathcal{F} \cup \mathcal{K} \to [0, 1]^{\infty}$ **to extend** d_{sub} **to** $\mathcal{F} \cup \mathcal{K}$**.**

The metrics d_{sub} and d_{cut}

Theorem (Borgs, Chayes, Lovász, Sós, Szegedy and Vesztergombi)

Let (G_n) be a sequence of graphs and let κ be a kernel. Then

$d_{\mathrm{sub}}(G_n, \kappa) \to 0$ **if and only if** $d_{\mathrm{cut}}(G_n, \kappa) \to 0$**.**

Corollary. Let κ and κ' be two kernels. Then $s(F, \kappa) = s(F, \kappa')$ **for every F if and only if** $d_{\mathrm{cut}}(\kappa, \kappa') = 0$**.**

Percolation on General Sequences

Theorem (B., Borgs, Chayes, Riordan)

Let (G_n) be a sequence of dense graphs with $|G_n| = n$, let λ_n be the largest eigenvalue of the adjacency matrix of G_n. For a constant $c > 0$,

set $\qquad p_n = c/\lambda_n.$ **$(< 1$ if n is large)**

(a) If $c \leq 1$, then all components of $G_n(p_n)$ are of size $o_p(n)$.

(b) If $c > 1$, then the the largest component of $G_n(p_n)$ has size $\Theta(n)$ whp.

VI. SPARSE GRAPHS

This is joint work with Oliver Riordan (Oxford).

Given a function $p = p(n)$ and a graph G with n vertices and $e(G) = \Theta(pn^2)$ edges, set

$$s_p(F, G) = \frac{\operatorname{emb}(F,G)}{p^{e(F)} \, (n)_{|F|}} = \frac{\operatorname{aut}(F) \, X_F(G)}{p^{e(F)} \, X_F(K_n)}.$$

Clearly,

$$s_p(F, G) = \frac{\operatorname{emb}(F,G)}{\mathbb{E}\Big(\operatorname{emb}(F,G(n,p))\Big)}.$$

If $0 < p < 1$ is constant, then we can define a map s as before, but now s maps $\mathcal{F}$ into the compact space $\prod_{F \in \mathcal{F}}[0, p^{-e(F)}]$.

However, if $p = p(n) = o(1)$ then in order to work in a compact space, we need that $s_p(F, G) \le c_F$ for all G and F we consider.

[The 'probability' p is often suppressed.]

Assumptions

Sequence (G_n)**, usually** $|G_n| = n$**; not all** n**.**

Assumption 1 (bounded subgraph counts). For each F**, the parameter** $s_p(F, G_n)$ **is bounded.**

Assumption 2 (exponentially bdd subgr. counts). There is a constant C **such that, for each fixed** F**, we have**

$$\limsup s_p(F, G_n) \leq C^{e(F)} \text{ **as** } n \to \infty.$$

(In the dense case these ineqs always hold.)

Kernels

As in BJR, a kernel is a symmetric, measurable (but not necessarily bounded) function

$$\kappa : \; [0, 1]^2 \longrightarrow \mathbb{R}_+.$$

VI.1. Almost Dense Graphs

Almost Dense Case: $p(n) = n^{-o(1)}$

Conjecture. (Main Conjecture)

Let $p = p(n) = n^{-o(1)}$**, and let** $C > 0$ **be constant. Suppose that** (G_n) **is a sequence of graphs with** $|G_n| = n$ **such that, for every** F**,** $s_p(F, G_n)$ **converges to some constant** $0 \le c_F \le C^{e(F)}$**. Then there is a bounded kernel** κ **such that** $c_F = s(F, \kappa)$ **for every** F**.**

The Uniform Case

By convexity, for p constant, only $\kappa = p$ a.e. satisfies $s_p(K_2, \kappa) = s_p(C_4, \kappa) = 1$.

Conjecture. (Basic Conjecture) Let $p = p(n) = n^{-o(1)}$**, and let** (G_n) **be a sequence of graphs with** $|G_n| = n$, $e(G_n) = p\binom{n}{2}$, $s_p(C_4, G_n) \longrightarrow 1$**, and** $\sup_n s_p(F, G_n) < \infty$ **for each** F **(needed!). Then** $s_p(F, G_n) \longrightarrow 1$ **for every** F**. (May even assume** $\exists C > 0$ **s.t.** $\limsup_n s_p(F, G_n) \leq C^{e(F)} \ \forall F$**.)**

Difficulties

We cannot even prove that there is a single triangle, let alone about $p^3 n^3/6$ **triangles. (!!)**

Many of the arguments from the dense case do not carry over. E.g., almost all, i.e., all but $o(n^2)$, pairs of vertices have about the right number of common neighbours.

In the dense case almost all (all but $o(pn^2) = o(n^2)$) edges are in the right number of triangles, and hence $t_p(K_3, G_n) \to 1$.

In the sparse case (with $o(n^2)$ edges), one should rule out the possibility that many of these pairs of vertices fall in the $o(n^2)$ set with too few common neighbours.

VII. Models for Metrics

Problem. Given a metric d on graphs, find a 'natural' family of random graph models with the following two properties:

(i) for each model, the sequence of random graphs (G_n) generated by the model is Cauchy with respect to d with probability 1, and

(ii) for any sequence (G_n) with $|G_n| = n$ that is Cauchy with respect to d, there is a model from the family such that, if we interleave (G_n) with a sequence of random graphs from the model, the resulting sequence is still a.s. Cauchy.

Answers?

For $d = d_{\mathrm{cut}}$ the answer is **yes** in the dense case, since (G_n) is Cauchy if and only if $d_{\mathrm{cut}}(G_n, \kappa) \to 0$ for some kernel κ, while the dense inhomogeneous random graphs $G(n, \kappa)$ converge to κ in d_{cut} with probability 1. Thus our family consists of essentially one model $G(n, \kappa)$ for each kernel κ.

In the sparse case: **no entirely satisfactory answer** for any of the natural metrics one should consider.

Assuming that $np \to \infty$, there is an **almost completely satisfactory answer** for d_{cut}: if we impose the Bounded Density Assumption (whatever that is!), the sparse inhomogeneous models $G_p(n, \kappa)$ answer this Question.

Answers?

In the extremely sparse case, with $p(n) = 1/n$, the model $G_{1/n}(n, \kappa)$ is **very unsatisfactory** for an arbitrary sequence of sparse graphs, since it produces graphs with essentially no cycles.

Very recently B., Janson and Riordan proposed a more general (and still rather natural) model.

VIII. Final Remarks

There is a rich theory of sparse inhomogeneous random graphs waiting to be explored.

The beginnings of such a theory can be found in the papers of B., Janson and Riordan in the very sparse case, and of Borgs, Chayes, Lovász, Sós, Szegedy and Vesztergombi in the dense case.

Task: build a theory encompassing both and everything in-between. B. and Riordan has made some progress in this direction.

This is unlikely to be easy: there are numerous unexpected difficulties and pitfalls, and much work has to be done even to arrive at concrete problems.