Percolation on graphs

Oliver Riordan

Mathematical Institute
University of Oxford

Analysis of Algorithms, Vienna, July 2010

Outline

1. Background and classical results
   - Classical results
2. Inhomogeneous models
3. Metrics and models

Graphs

In this talk, all graphs are finite. Formally, $G = (V, E)$, where V is any finite set, and E a set of (unordered) pairs from V. Usually $V = \{1, 2, \ldots, n\}$.

Graph encodes one bit of information per pair of vertices, typically some notion of connectedness.

Selected examples

Vertices	Edge if
Atoms in crystal	Bond between them
Computers	Directly connected
People	Know each other
Web pages	Link from one to another
Chemicals in cell	React

In most cases random graphs are natural models.

Random graphs: classical models

Fix the vertex set $[n] = \{1, 2, \ldots, n\}$.

- $G(n, p)$: include each possible edge with probability p, independently. (Gilbert, 1959)
- $G(n, M)$: pick a graph with M edges, all such graphs equally likely. (Erdős and Rényi, 1959)

Often equivalent, with $p \sim M/\binom{n}{2}$.

Systematic study started by Erdős and Rényi; both models often named after them.

p can be constant or function of n; here appropriate scaling is $p = c/n$.

Small-world phenomenon

Many real-world networks show small diameter, or typical distances – logarithmic in number of vertices.
E.g. '6 degrees of separation'.

Watts and Strogatz (1998) proposed model with this feature.
$G(n, p)$ has it also!

In fact, even small amount of 'global randomness' gives small diameter; e.g., Bollobás and Chung (1988).

Message: absence of small-world would need explanation.

Percolation properties

We can ask many questions about graphs. Some of most interesting concern 'global connectivity' properties.

- What is the diameter?
- Are most or all vertices connected? How big is the largest connected part?
- How much remains connected if edges fail randomly?

Phase transition in $G(n,p)$: $p = c/n$

Classical result of Erdős and Rényi (1960).
Let $C_1(G)$ denote the number of vertices in the largest component.

Theorem

Let c be a constant, and let $G = G(n, c/n)$.
If $c < 1$ then there is a constant $A > 0$ such that
$$C_1(G) \leq A \log n \text{ whp (with high probability).}$$
If $c > 1$ then there is a constant ρ such that $C_1(G)/n \xrightarrow{\mathrm{p}} \rho$.
If $c = 1$ then $C_1(G)$ is of order $n^{2/3}$.

Near the critical value

What about other functions $p(n)$?

Can suppose $np \to c$.

- For $c \neq 1$, result unchanged.
- For $c = 1$, depends on how quickly.

Results proved by Bollobás 1984, Łuczak 1990,...., Janson, Knuth, Łuczak and Pittel 1993,...

Here interested in 'weak' results, but for more general models.

Branching processes

There is a natural local approximation to $G(n, c/n)$.

The Galton–Watson process $\mathfrak{X}_c$

- Start with one individual.
- Each individual has random no. of children with a Poisson Po(c) distribution.
- Numbers independent.

Survival probability

Let $\rho(c) = \mathbb{P}(\ \mathfrak{X}_c$ survives forever).
Observation:

$$\rho(c) = 1 - e^{-c\rho(c)}$$

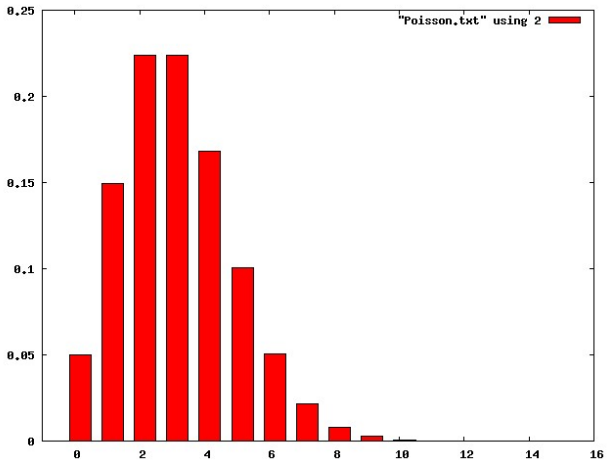
Proof:

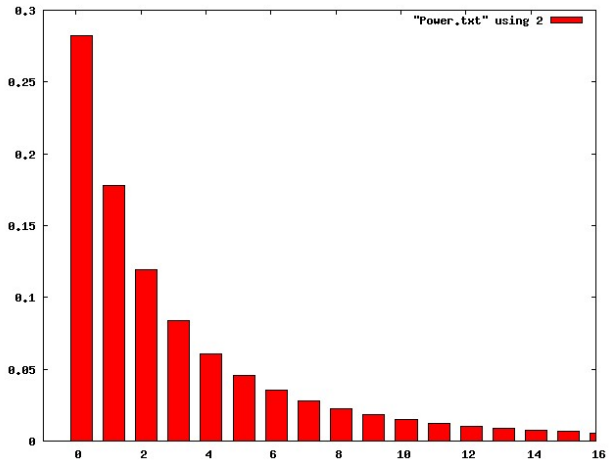
- each individual in generation 1 survives with probability $\rho(c)$.
- Number that survive $\mathrm{Po}(c\rho(c))$.
- Process survives if at least one does.

Does not quite determine $\rho(c)$: 0 also a solution.

Degree distribution

- In $G(n, c/n)$, degree distribution asymptotically Poisson.





- Many real-world examples asymptotically power-law. (Barabási–Albert, Faloutsos[3], KKRRT 1999)
- Suggests need for *inhomogeneous* models.

‘Scale-free’ models.

Barabási and Albert (BA) (1999)
Kumar, Raghavan, Rajagopalan, Sivakumar, Tomkins and Upfal (2000)
Aiello–Chung–Lu (2000)
Linearized Chord Diagram model (Bollobás, R (2004-))
Buckley and Osthus (2004)
Cooper and Frieze (2003)
...

Problems with models

Two main problems:

- Many models hard to analyze. (But LCD ok.)
- Analyzing models one by one is a never-ending process!

Bollobás–Janson–Riordan model

Outline definition:

- Vertices have "types" in some set $\mathcal{S}$.
- Usually $\mathcal{S} = [0, 1]$; can simply be finite.
- Edge probabilities depend on both types, leading to
- *kernels* κ: symmetric functions $\kappa : [0, 1]^2 \to [0, \infty)$.
- Types are random; given types (x_i), edges independent, with probabilities $\kappa(x_i, x_j)/n$.

The finite-type case is "folklore" as a model.

Phase transition

The definition makes the BJR model amenable to analysis.

Key object: multi-type Poisson branching process $\mathfrak{X}_\kappa$.

Survival probability $\rho(x)$ satisfies

$$\rho = 1 - e^{-T_\kappa \rho},$$

where T_κ is the operator defined by

$$(T_\kappa(f))(x) = \int_0^1 \kappa(x,y) f(y)\,\mathrm{d}y.$$

Phase transition

Theorem

Let $\kappa : [0,1]^2 \to [0,\infty)$ be a kernel, and let $G_n = G(n,\kappa)$ denote the random graph constructed by the BJR model.

(i) *If $\|T_\kappa\| \leq 1$, then $C_1(G_n) = o_p(n)$.*

(ii) *If $\|T_\kappa\| > 1$, then $C_1(G_n) = \Theta(n)$ whp.*

If κ is irreducible, then $C_1(G_n)/n \xrightarrow{p} \rho(\kappa)$ and $C_2(G_n) = o_p(n)$.

To study transition: parametrize. I.e., study function $c \mapsto \rho(c\kappa)$ with κ constant.

Examples

$\kappa = 1$: classical model $G(n, c/n)$.

$\kappa(x, y) = \psi(x)\psi(y)$: 'rank 1' case. Studied by (e.g.) Chung and Lu, Norros and Reittu. Range of behaviour depending on which norms of ψ are finite.

$\kappa(x, y) = 1/\sqrt{xy}$: approximation to Barabási–Albert model.

$\kappa(x, y) = 1/\max\{x, y\}$ (-1): CHKNS model or Turova's model. Infinite order phase transision.

Susceptibility

Branching process analysis yields many other properties. E.g., susceptibility (Janson and R. 2010+):

Theorem

In the subcritical case,

$$\chi(\kappa) = \langle (I - T_\kappa)^{-1} 1, 1 \rangle.$$

For supercritical case, use duality. (Janson and R. 2010+)

Diameter

Bollobás, Janson and R. (2007) studied diameter (in certain cases).

Result new for $G(n, c/n)$! (See also Fernholz and Ramachandran.)

Previously, results for subcritical case by Łuczak (1998); strongly supercritical by Bollobás. Some partial results in between by Chung and Lu (2001).

Very precise results for entire supercritical range by R. and Wormald (2010+).

Diameter of $G(n, p)$

Theorem

Let $\lambda > 1$ be fixed, and define $\lambda^* < 1$ by $\lambda^* e^{-\lambda^*} = \lambda e^{-\lambda}$. Then

$$\mathrm{diam}(G(n, \lambda/n)) = \frac{\log n}{\log \lambda} + 2\frac{\log n}{\log(1/\lambda^*)} + O_{\mathrm{p}}(1).$$

Same with $\lambda \to \infty$, with 2-point concentration.

Some version of this announced by Bruce Reed; not yet written up.

Diameter of $G(n,p)$

Theorem

Let $\varepsilon = \varepsilon(n)$ (bounded) satisfy $\varepsilon^3 n \to \infty$. Set $\lambda = \lambda(n) = 1 + \varepsilon$, and define $\lambda^* < 1$ by $\lambda^* e^{-\lambda^*} = \lambda e^{-\lambda}$. Then

$$\mathrm{diam}(G(n,\lambda/n)) = \frac{\log(\varepsilon^3 n)}{\log \lambda} + 2\frac{\log(\varepsilon^3 n)}{\log(1/\lambda^*)} + O_{\mathrm{p}}(1/\varepsilon).$$

Independent weaker result by Ding, Kim, Lubetzky and Peres.

Inside the window: Addario-Berry, Broutin and Goldschmidt.

Distribution of the error term

Let $z_1 > z_2 > \cdots$ list the points of a Poisson process $\mathcal{P}$ on $\mathbb{R}$ with density function $f(x) = 4\gamma_1 e^{-x}$ in decreasing order. For each $1 \le i < j$, let T_{ij} be a random variable with $\mathbb{P}(T_{ij} > x) = \exp(-e^x)$, independent of each other and of $\mathcal{P}$. Let $D = \sup\{z_i + z_j + T_{ij}\}$.

Theorem

Let $\varepsilon = \varepsilon(n) > 0$ satisfy $\varepsilon \to 0$ and $\varepsilon^3 n \to \infty$, and let $\lambda = 1 + \varepsilon$. For any constant c

$$\mathbb{P}\left(\mathrm{diam}(G(n,\lambda/n)) \ge \frac{\log(\varepsilon^3 n)}{\log \lambda} + 2\frac{\log(\varepsilon^3 n)}{\log(1/\lambda^*)} + c/\varepsilon\right)$$

converges to $\mathbb{P}(D \ge c)$ as $n \to \infty$.

Summary

BJR model unifies many earlier sparse models.

Connection to a continuous object makes analysis possible.

Key tool: branching processes.

Methods feed back to give new results even for $G(n, p)$.

Extensions....

BJR model with clustering. (Bollobás, Janson, R. 2010)

Allows 'tuning' clustering coefficient, mixing coefficient, ...

Analysis involves multi-type compound Poisson branching processes; operator T becomes non-linear.

The need for metrics

A fundamental question is how to tell whether a (random or deterministic) model graph is close to the real-world example.

- Can consider particular parameters, but not satisfactory.
- Typically *not* a sampling question: there is only one world-wide web.

The edit distance

Distance 0: isomorphic.
Distance 1: isomorphic after changing one edge.

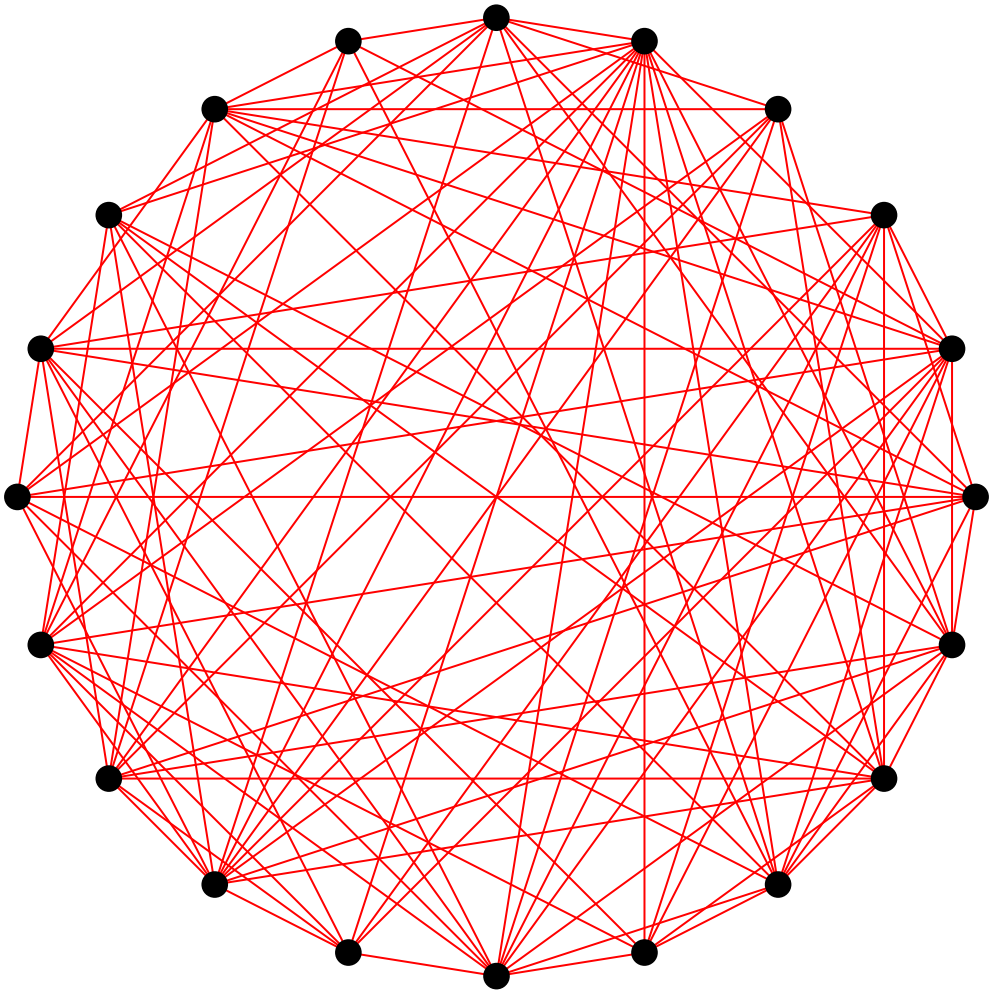
The edit distance

If $|G| = |H| = n$ then

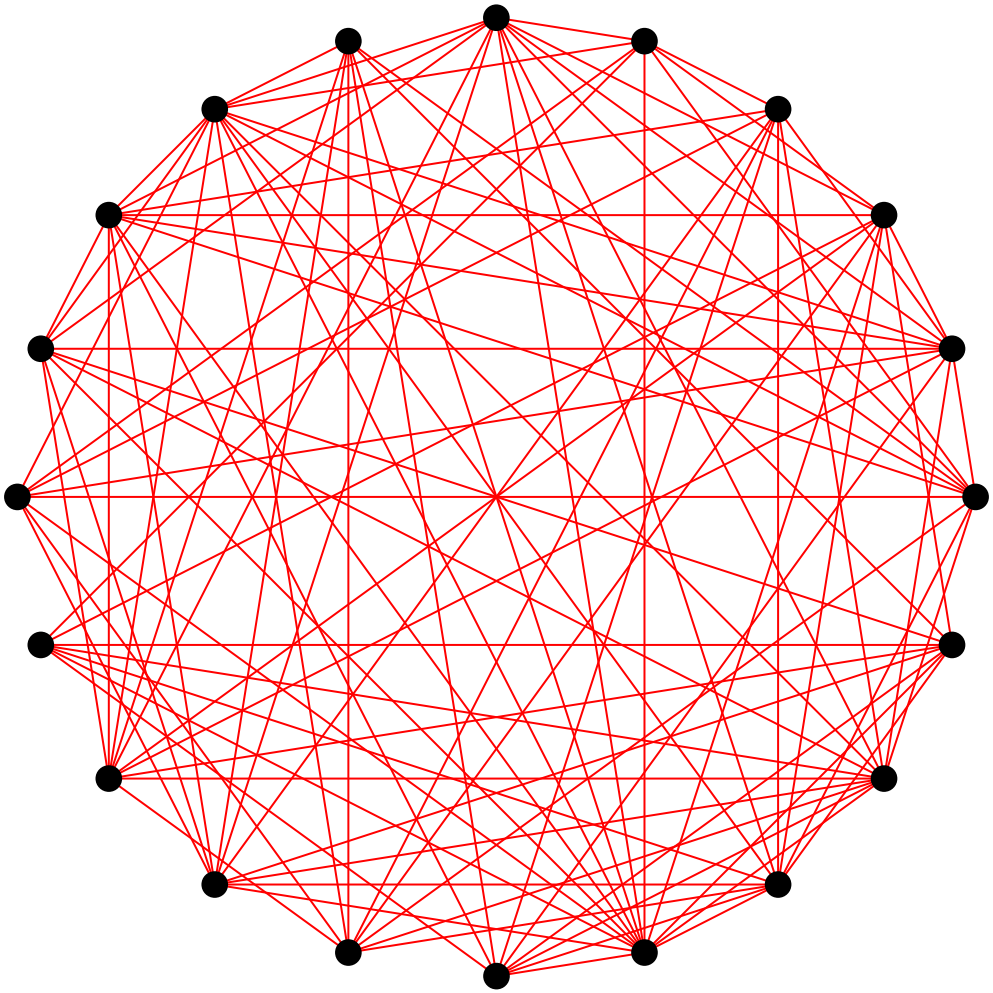
$$d_{\text{edit}}(G, H) = \min\{\, |E(G) \,\Delta\, E(H')| : H' \cong H \,\}.$$

- Explicitly defined by Axenovich, Kézdy and Martin (2008).
- Used by Erdős (1966) and Simonovits (1966).

Two instances of $G(n, 1/2)$



Two instances of $G(n, 1/2)$



Two instances of $G(n, 1/2)$

In edit distance, far apart. However...

There is a very strong intuitive sense in which two 'typical' instances of $G(n, 1/2)$ are 'similar'.

- For 'most' properties, probability close to 0 or 1.
- Typical numerical properties close: e.g., # triangles.

Goes back to Erdős and Rényi.

The subgraph distance

The *density* of F in G is

$$s(F,G) = \frac{\mathrm{emb}(F,G)}{\mathrm{emb}(F,K_n)} = \frac{\mathrm{emb}(F,G)}{n(n-1)\dots(n-k+1)}$$

where $n = |G|$.

Definition

The *subgraph metric* may be defined by

$$d_{\mathrm{sub}}(G,H) = \sum_F 2^{-|F|^2} |s(F,G) - s(F,H)|.$$

Usually only care about the topology: (G_n) is Cauchy in d_{sub} iff $s(F,G_n)$ converges for all F.

Limit points

Every sequence has a subsequence which is Cauchy, so completion is compact. Is there a nice description?

Theorem (Lovász and Szegedy (2006))

Given numbers s_F, there is a sequence (G_n) with $|G_n| \to \infty$ such that $s(F, G_n) \to s_F$ for all F if and only if there is a kernel $\kappa : [0,1]^2 \to [0,1]$ such that

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

for all F.

(We take vertex set of F to be $[k]$ where $k = |F|$.)

Global similarity

Borgs, Chayes, Lovász, Sós and Vesztergombi (2008) defined a metric, the *cut metric* that measures global similarity.

Roughly speaking, two graphs are close if can overlay them so that inside each large set, or across each large cut, almost the same *number* of edges in each..

Again limits are kernels; moreover, equivalent to subgraph metric!

Random subgraphs of arbitrary graphs

Can view $G(n,p)$ as $K_n(p)$: random subgraph of K_n.
Percolation threshold is $p^*(n) = 1/n$.
What is threshold in $H_n(p_n)$?

Theorem (Bollobás, Borgs, Chayes, R. (2010))

Suppose H_n has n vertices and $\Theta(n^2)$ edges. Let $\lambda_n =$ largest eigenvalue of H_n, let $c > 0$ be constant and set $G_n = H_n(c/\lambda_n)$.

- *If $c \leq 1$, then $C_1(G_n) = o_p(n)$.*
- *If $c > 1$ then $C_1(G_n) = \Theta(n)$ whp.*

I.e., threshold is $1/\lambda_n$; no assumptions on (H_n).

Size of giant component

Certainly need to restrict (H_n) to get a result; natural assumption is convergence in $\delta_\square$.

Theorem

Suppose $|H_n| = n$ *and* $\delta_\square(H_n, \kappa) \to 0$. *With* c *constant set* $G_n = H_n(c/n)$.

(i) *If* $c \le \|T_\kappa\|^{-1}$, *then* $C_1(G_n) = o_{\mathrm{p}}(n)$.

(ii) *If* $c > \|T_\kappa\|^{-1}$, *then* $C_1(G_n) = \Theta(n)$ *whp.*

If κ *is irreducible, then* $C_1(G_n)/n \xrightarrow{\mathrm{p}} \rho(c\kappa)$ *and* $C_2(G_n) = o_{\mathrm{p}}(n)$.

Implies previous result by (sub)subsequences.

Combining the results

Can represent a graph with independence between edges by matrix A_n of edge probabilities. In fact, take probabilities as a_{ij}/n; constant matrix gives $G(n, c/n)$.

BJR result and BBCR result establish threshold assuming convergence to κ.

BJR: κ arbitrary, strong (L^1) notion of convergence.

BBCR: only need cut convergence, but κ and A_n bounded.

Bollobás, Janson, R. (2010+)

Need only cut convergence; no boundedness assumptions.

Metrics and models

In dense case, natural random graph model for each (bounded!) kernel, introduced by Lovász and Szegedy (2006).

Since *any* sequence has subsequence converging to a kernel, this family of models is in some sense complete: *any* graph is 'close' to one of these random graphs.

The sparse case

Main question

Is there a family of sparse random graph models that 'covers' the space of sparse graphs?

Many difficulties: which metric should we use? Are $\delta_\square$ and d_{sub} still equivalent?

Many partial results and open questions:
Bollobás and R., Metrics for sparse graphs, 2009,
Bollobás and R., Sparse graphs: metrics and random models. (On arXiv).