ADFOCS 2006 Saarbrücken

The Erdős-Rényi Phase Transition

Joel Spencer

TP! trivial being! I have received your letter, you should have written already a week ago.

The spirit of Cantor was with me for some length of time during the last few days, the results of our encounters are the following . . .

letter, Paul Erdős to Paul Turán

November 11, 1936

Paul Erdős and Alfred Rényi

On the Evolution of Random Graphs

Magyar Tud. Akad. Mat. Kutató Int. Közl

volume 8, 17-61, 1960

$\Gamma_{n,N(n)}$: n vertices, random $N(n)$ edges

[...] the largest component of $\Gamma_{n,N(n)}$ is of order $\log n$ for $\frac{N(n)}{n} \sim c < \frac{1}{2}$, of order $n^{2/3}$ for $\frac{N(n)}{n} \sim \frac{1}{2}$ and of order n for $\frac{N(n)}{n} \sim c > \frac{1}{2}$. This double "jump" when c passes the value $\frac{1}{2}$ is one of the most striking facts concerning random graphs.

The "Double Jump"

$G(n,p)$, $p = \frac{c}{n}$ (or $\sim \frac{c}{2}n$ edges)

(Average Degree c, "natural" model)

- $c < 1$

  Biggest Component $O(\ln n)$

  $|C_1| \sim |C_2| \sim \ldots$

  All Components simple (= tree/unicyclic)

- $c = 1$

  Biggest Component $\Theta(n^{2/3})$

  $|C_1| n^{-2/3}$ nontrivial distribution

  $|C_2|/|C_1|$ nontrivial distribution

  Complexity of C_1 nontrivial distribution

- $c > 1$

  Giant Component $|C_1| \sim yn$, $y = y(c) > 0$

  All other $|C_i| = O(\ln n)$ and simple

Galton-Watson Birth Process

Root node "Eve"

Each node has $Po(c)$ children

(Poisson: $\Pr[Po(c) = k] = e^{-c}c^k/k!$)

$T = T_c$ is total size

- $c < 1$

  T finite

- $c = 1$

  T finite

  $E[T]$ infinite (heavy tail)

- $c > 1$

  $\Pr[T = \infty] = y = y(c) > 0$

Galton-Watson Exact

$$\Pr[T_c = u] = \frac{e^{-uc}(uc)^{u-1}}{u!}$$

$$\Pr[T_1 = u] = \frac{e^{-u}u^{u-1}}{u!} = \Theta(u^{-3/2})$$

For $c > 1$, $\Pr[T = \infty] = y = y(c) > 0$ where

$$1 - y = e^{-cy}$$

For $c < 1$, $\alpha := ce^{-c} < 1$

$\Pr[T_c > u] = O(\alpha^u)$ Exponential Tail

Duality: $d < 1 < c$ with $de^{-d} = c^{-c}$

Conditioning on

$Po(c)$ process being finite

gives the $Po(d)$ process

Math Physics Bond Percolation

Z^d. Bond "open" with probability p

There exists a critical probability p_c

• Subcritical, $p < p_c$.

All C finite, $E[|C(\vec{0})|]$ finite

$\Pr[|C(\vec{0})| \geq u]$ exponential tail

• Supercritical, $p > p_c$.

Unique Infinite Component

$E[|C(\vec{0})|]$ infinite

$\Pr[|C(\vec{0})| \geq u | \text{finite}]$ exponential tail

• Critical, $p = p_c$.

All C finite, $E[|C(\vec{0})|]$ infinite, heavy tail

Random 3-SAT

n Boolean $x_1, \ldots, x_n$

$L = \{x_1, \overline{x_1}, \ldots, x_n, \overline{x_n}\}$ literals

Random Clauses $C_i = y_{i1} \vee y_{i2} \vee y_{i3}$, $y_{ij} \in L$

$f(m) :=$ Pr[$C_1 \wedge \cdots \wedge C_m$satisfiable]

Conjecture: There exists critical c_0

- Subcritical, $c < c_0$, $f(cn) \sim 1$
- Supercritical, $c > c_0$, $f(cn) \sim 0$

Friedgut: Criticality, but possibly nonuniform

Evolution of n-Cube

Ajtai, Komlos, Szemeredi

Bollobas, Luczak, Kohayakawa

Borgs, Chayes, Slade, JS, van der Hofstad

$p = c/n$

$c < 1$ subcritical

$c > 1$ giant $\Omega(2^n)$ component

Much more!

The Critical Window $p = \frac{1}{n} + \lambda n^{-4/3}$

- $\lambda(n) \to -\infty$ Subcritical

  Biggest Component $o(n^{2/3})$

  $|C_1| \sim |C_2| \sim \ldots$

  All Components simple

- λ constant. The Critical Window

  Biggest Component $\Theta(n^{2/3})$

  $|C_1| n^{-2/3}$ nontrivial distribution

  $|C_2|/|C_1|$ nontrivial distribution

  Complexity of C_1 nontrivial distribution

- $\lambda(n) \to +\infty$ Supercritical

  Dominant Component $|C_1| \gg n^{2/3}$

  High Complexity

  All other $|C_i| = o(n^{2/3})$ and simple

What is the Critical Window?

Difficult in General

When Dominant Component is Emerging

Subcritical: Biggest about same size

Supercritical: Biggest $\gg$ second

Susceptibility $\chi(G) = E[|C(0)|] = \frac{1}{n}\sum |C_i|^2$

Largest Component starts to dominate

Subcritical: $\frac{1}{n}|C_1|^2 \ll \chi(G)$

Critical: $\frac{1}{n}|C_1|^2 = O(\chi(G))$

Supercritical: $\frac{1}{n}|C_1|^2 \sim \chi(G)$

Computer Experiment (Try It!)

$n = 50000$ vertices. Start: Empty

Add random edges

Parametrize $e/\binom{n}{2} = (1 + \lambda n^{-1/3})/n$

Merge-Find for Component Size/Complexity

$-4 \leq \lambda \leq +4$, $|C_i| = c_i n^{2/3}$

See biggest merge into dominant

A Strange Physics

Components $c_i n^{2/3}$, $c_j n^{2/3}$

$\lambda \leftarrow \lambda + d\lambda$, $p \leftarrow p + n^{-4/3} d\lambda$

Merge with probability $c_i c_j d\lambda$

Increment Complexity $\frac{1}{2} c_i^2 d\lambda$

An Open Question

What is the *critical window* for random k-SAT

That is, can you find a parametrization

$$m = f_0(n) + \lambda f_1(n)$$

so that

- Subcritical $\lambda \to -\infty$, Pr[SAT] $\to 1$
- Supercritical $\lambda \to +\infty$, Pr[SAT] $\to 0$
- Critical Window λ fixed, Pr[SAT] $\to g(\lambda)$

where

$$\lim_{\lambda \to -\infty} g(\lambda) = 1 \text{ and } \lim_{\lambda \to +\infty} g(\lambda) = 0$$

(*Maybe* $m = c_0 n + \lambda n^{1-\beta}$ for some "critical exponent" β.)

Galton-Watson Near Criticality

$\Pr[T_1 \geq u] \sim cu^{-1/2}$

$\Pr[T_{1+\epsilon} = \infty] \sim 2\epsilon$

Conditioning on finite, $T_{1+\epsilon}$ becomes $T_{1-\epsilon+o(\epsilon)}$

$\Pr[T_{1-\epsilon} \geq u] \sim \Pr[\infty > T_{1+\epsilon} \geq u]$

If $u = o(\epsilon^{-2})$ (can't see ϵ):

$\Pr[\infty > T_{1+\epsilon} \geq u] \sim \Pr[T_1 \geq u] \sim cu^{-1/2}$

If $u = \Theta(\epsilon^{-2})$ (somewhat see ϵ):

$\Pr[\infty > T_{1+\epsilon} \geq u] = \Theta(\Pr[T_1 \geq u]) = \Theta(u^{-1/2})$

If $u \gg \epsilon^{-2}$ (strong ϵ effect):

$\Pr[\infty > T_{1+\epsilon} \geq u] \sim \Pr[T_1 \geq u]e^{-u\epsilon^2/2}$

$$\begin{aligned} \frac{\Pr[T_{1\pm\epsilon} = u]}{\Pr[T_1 = u]} &= [e^{\mp\epsilon}]^u(1 \pm \epsilon)^{u-1} \\ &\sim [(1 \pm \epsilon)e^{\mp\epsilon}]^u \\ &= e^{-(1+o(1))u\epsilon^2/2} \end{aligned}$$

Galton-Watson as Walk

$Z_i \sim Po(c)$, $i = 1, 2, \ldots$

$Y_0 = 1$ (Eve)

$Y_i = Y_{i-1} + Z_i - 1$ (Z_i children and dies)

Fictitious Continuation

$T =$ min t with $Y_t = 0$

(If no such t, $T = \infty$)

$c < 1$ negative drift, T finite

$c > 1$ positive drift, maybe T infinite

$c = 1$ zero drift, delicate

$C(v)$ in $G(n,p)$ as BFS Walk

$Y_0 = 1$ (Root v)

$Y_i = Y_{i-1} + Z_i - 1$ (pop queue/add Z_i new)

where $Z_i \sim \mathrm{BIN}[n-(i-1)-Y_{i-1}, p]$

The Link:

When $p \sim \frac{c}{n}$

and $i - 1 + Y_{i-1} = o(n)$

Z_i is roughly $Po(c)$

$|C(v)|$, T_c similar while small

Ecological Limitation: Success in BFS in $G(n,p)$

is selflimiting. "Eating your seed corn"

Rough (but Accurate) Link

$p = \frac{c}{n}$, $c > 1$

$C(v)$ like T_c if finite

With probability y, T_c infinite

Corresponding $C(v)$ become large

All merge to form giant $\sim yn$ component

$p = \frac{1+\epsilon}{n}$, $o(1) = \epsilon \gg n^{-1/3}$

With probability $\sim 2\epsilon$, T_c infinite

Corresponding $C(v)$ become large

All merge to form dominant $\sim 2\epsilon n$ component

Finite T_c have small $|C(v)| < \epsilon^{-2+}$

$\epsilon \gg n^{-1/3}$ small/dominant dichotomy

The (easy!) subcritical case

$G \sim G(n,p)$, $p = \frac{c}{n}$, $c < 1$

$|C(v)|$ dominated by Galton Watson T_c

$\Pr[T_c > u] < \alpha^u = o(n^{-1})$ for $u = K \ln n$

Therefore:

$$\text{NO } |C(v)| > K \ln n$$

Why $\Theta(n^{2/3})$ at $p = \frac{1}{n}$

Ignore Ecological Limitation (so rough!)

$\Pr[|C(v)| \geq u] \sim \Pr[T_1 \geq u] = \Theta(u^{-1/2})$

$X_u :=$ number v with $|C(v)| \geq u$

$E[X_u] = \Theta(nu^{-1/2})$

$X_u \neq 0 \Rightarrow X_u \geq u$

$\Pr[X_u \neq 0] = O(nu^{-3/2}) = O(1)$ when $u = \Theta(n^{2/3})$

BFS on $G(n,p)$

Root 0, Nonroots $1,\ldots,n-1$

$T_j^* = i$: Vertex j joins BFS tree at i-th opportunity (*fictitious continuation!*) T_j^* geometric

If $X=k$ *then* precisely $k-1$ of $T_j \leq k$

$A_1 :=$ Pr[BIN$[n-1, 1-(1-p)^k] = k-1$

Take $p \sim \frac{c}{n}$ with $c > 1$

A_1 very small unless $k = O(\ln n)$ *or* $k \sim yn$ with $1-e^{-cy} = y$

Thus: All components small or giant

The Giant Exists and is Unique!

$t = O(\ln n)$ same as Galton-Watson $\Rightarrow$

$$\Pr[|C(v)| = O(\ln n)] \sim \Pr[T_c < \infty] = 1 - y$$

Karp Approach: Keep generating components

After $O(1)$ tries (still $\sim n$ reservoir) get giant

Now $n' = n(1-y)$, $p = d/n'$, $d < 1 < c$

Now *subcritical*, no more giants!

Duality: $G(n, c/n)$ minus giant component is like $G(n', d/n')$ (c, d conjugate)

BFS on $G(n,p)$ conditioned

Condition: Precisely $k-1$ of $T_j^* \leq k$

WLOG $T_j^* \leq k$ for $1 \leq j \leq k-1$

$T_j^* \rightarrow T_j$, truncated geometric

$\Pr[T_j = u] = \frac{p(1-p)^{u-1}}{1-(1-p)^k}$

$Z_t :=$ number of $T_j = t$ (join queue at time t)

$Y_0 = 1$, $Y_t = Y_{t-1} - 1 + Z_t$ (queue size)

TREE: $Y_t > 0$ for $1 \leq t < k$.

$\Pr[|C(0)| = k] = A_1 \Pr[\text{TREE}]$

Example: Connected on $0, 1, 2, 3, 4, 5$

1	2	3	4	5
N	N	Y	Y	N
N	N	-	-	N
Y	N	-	-	N
-	Y	-	-	Y
-	-	-	-	-
-	-	-	-	-

$T_3 = T_4 = 1$, $T_1 = 3$, $T_2 = T_5 = 4$

A_1: All $T_j \leq 6$

$\vec{Z} = (2, 0, 1, 2, 0, 0)$

Walk $\vec{Y} = (1, 2, 1, 1, 2, 1, 0)$

TREE: BFS doesn't terminate early

Tree Edges $03, 04, 41, 12, 15$

Pr[TREE] with $p \sim \frac{c}{k}$

Left Z_i Poisson $\frac{c}{1-e^{-c}}$

Galton-Watson Pr[ESC] $\sim 1 - e^{-c}$

Right $Z_i^* = Z_{k-i}; Y_i^* = Y_{k-i}$

$Y_0^* = 0$, $Y_i^* = Y_{i-1}^* + 1 - Z_i^*$

Z_i^* Poisson $\frac{ce^{-c}}{1-e^{-c}}$

Pr[ESC*] $\sim 1 - \frac{ce^{-c}}{1-e^{-c}}$

Chernoff: $Y_i > 0$ in middle

Pr[TREE] $\sim$ Pr[ESC] Pr[ESC^*] $\rightarrow 1-(c+1)e^{-c}$

It is six in the morning.

The house is asleep.

Nice music is playing.

I prove and conjecture.

– Paul Erdős, in letter to Vera Sós