from λ to $\lambda + d\lambda$ there is a probability $c_1c_2d\lambda$ that they will merge. Components have a peculiar gravitation in which the probability of merging is proportional to their sizes. With probability $(c_1^2/2)d\lambda$ there will be a new internal edge in a component of size $c_1n^{2/3}$ so that large components rarely remain trees. Simultaneously, big components are eating up other vertices.

With $\lambda = -10^6$, say, we have feudalism. Many small components (castles) are each vying to be the largest. As λ increases the components increase in size and a few large components (nations) emerge. An already large France has much better chances of becoming larger than a smaller Andorra. The largest components tend strongly to merge and by $\lambda = +10^6$ it is very likely that a giant component, Roman Empire, has emerged. With high probability this component is nevermore challenged for supremacy but continues absorbing smaller components until full connectivity - One World - is achieved.

An Continuous Model. In discussions at St. Flour it became apparent that there was a continuous model underlying the asymptotic behavior of $G(n,p)$ with $p = n^{-1} + \lambda n^{-4/3}$. The following should be regarded as only tentative steps toward defining of that continuous model. For fixed λ and k arbitrarily large but fixed one can look at the k largest components of $G(n,p)$ and parametrize them $x_1n^{2/3}, \ldots, x_kn^{2/3}$ in decreasing order. One can give explicitly a limiting distribution function $H(x_1, \ldots, x_k)$ for these values. Now one can go to the limit with k and consider the "state" $P(\lambda)$ at "time" λ to be an infinite sequence $x_1 > x_2 > \ldots$ of decreasing reals. There will be a distribution over the possible sequences. The sequences must be well-behaved; one can show, for example, that the number of x_i bigger than c must be asymptotic to $\frac{2}{3}(2\pi)^{-1/2}c^{-3/2}$ as $c \to 0$. (There is further information concerning the nature of the components - e.g., are they trees, unicyclic,...- that could also be added.) Now the intriguing thing is the "gravity" that defines $P(\lambda + d\lambda)$ in terms of $P(\lambda)$ in an appropriate limiting sense. If $P(\lambda)$ has terms x_i, x_j then with probability $x_ix_jd\lambda$ they will "merge" and form a single term with value $x_i + x_j$. This corresponds to certain coagulation models in physics though in the physical world the probability of coagulation depends on the surface area (and perhaps other invariants) of the objects) while here it depends only on their sizes. So it seems there should be a probability space whose elements are histories - i.e., the value of $P(\lambda)$ for all real λ - where the change from $P(\lambda)$ to $P(\lambda + d\lambda)$ is governed by these coagulation laws and where further there have to be some appropriate entry laws so that each $P(\lambda)$ has the appropriate distribution. Not that any of this has been done - but in theory there is a theory!

with $k-1+l$ edges, where c_l was given by a specific recurrence. Asymptotically in l,

$$c_l \sim \left(\frac{e}{12l}(1+o(1))\right)^{l/2}.$$

The calculation for $X^{(l)}$, the number of such components on k vertices, leads to extra factors of $c_l k^{\frac{3}{2}l}$ and n^{-l} which gives $c_l c^{\frac{3}{2}l}$. For fixed a, b, λ, l the number $X^{(l)}$ of components of size between $an^{2/3}$ and $bn^{2/3}$ with $l-1$ more edges than vertices satisfies

$$\lim_{n\to\infty} E[X^{(l)}] = \int_a^b e^{-\frac{c^3}{6}-\frac{\lambda^2 c}{2}+\frac{\lambda c^2}{2}} c^{-5/2}(2\pi)^{-1/2}(c_l c^{\frac{3}{2}l})dc$$

and letting X^* be the total number of components of size between $an^{2/3}$ and $bn^{2/3}$

$$\lim_{n\to\infty} E[X^*] = \int_a^b e^{-\frac{c^3}{6}-\frac{\lambda^2 c}{2}+\frac{\lambda c^2}{2}} c^{-5/2}(2\pi)^{-1/2} g(c)dc$$

where

$$g(c) = \sum_{l=0}^{\infty} c_l c^{\frac{3}{2}l}$$

a sum convergent for all c, (here $c_0 = 1$). A component of size $\sim cn^{2/3}$ will have probability $c_l c^{\frac{3}{2}l}/g(c)$ of having $l-1$ more edges than vertices, independent of λ. As $\lim_{c\to 0} g(c) = 1$, most components of size $\epsilon n^{2/3}$, $\epsilon << 1$, are trees but as c gets bigger the distribution on l moves inexoribly higher.

An Overview. For any fixed λ the sizes of the largest components are of the form $cn^{2/3}$ with a distribution over the constant. For $\lambda = -10^6$ there is some positive limiting probability that the largest component is bigger than $10^6 n^{2/3}$ and for $\lambda = +10^6$ there is some positive limiting probability that the largest component is smaller than $10^{-6}n^{2/3}$, though both these probabilities are minuscule. The functions integrated have a pole at $c = 0$, reflecting the notion that for any λ there should be many components of size near $\epsilon n^{2/3}$ for $\epsilon = \epsilon(\lambda)$ appropriately small. When λ is large negative (e.g., -10^6) the largest component is likely to be $\epsilon n^{2/3}$, ϵ small, and there will be many components of nearly that size. The nontree components will be a negligible fraction of the tree components.

Now consider the evolution of $G(n,p)$ in terms of λ. Suppose that at a given λ there are components of size $c_1 n^{2/3}$ and $c_2 n^{2/3}$. When we move so that

$$\sum_{i=1}^{k-1} -\ln(1-\frac{i}{n}) = \frac{k^2}{2n} + \frac{k^3}{6n^2} + o(1) = \frac{k^2}{2n} + \frac{c^3}{6} + o(1)$$

Also

$$p^{k-1} = n^{1-k}(1+\frac{\lambda}{n^{1/3}})^{k-1}$$

$$(k-1)\ln(1+\frac{\lambda}{n^{1/3}}) = (k-1)(\frac{\lambda}{n^{1/3}} - \frac{\lambda^2}{2n^{2/3}} + O(n^{-1})) = \frac{\lambda k}{n^{1/3}} - \frac{\lambda^2 c}{2} + o(1)$$

Also

$$\ln(1-p) = -p + O(n^{-2}) = -\frac{1}{n} - \frac{\lambda}{n^{4/3}} + O(n^{-2})$$

and

$$k(n-k) + \binom{k}{2} - (k-1) = kn - \frac{k^2}{2} + O(n^{2/3})$$

so that

$$[k(n-k) + \binom{k}{2} - (k-1)]\ln(1-p) = -k + \frac{k^2}{2n} - \frac{\lambda k}{n^{1/3}} + \frac{\lambda c^2}{2} + o(1)$$

and

$$E[X] \sim \frac{n^k k^{k-2}}{k^k\sqrt{2\pi k}n^{k-1}}e^A$$

where

$$\begin{aligned} A &= k - \tfrac{k^2}{2n} - \tfrac{c^3}{6} + \tfrac{\lambda k}{n^{1/3}} - \tfrac{\lambda^2 c}{2} - k + \tfrac{k^2}{2n} - \tfrac{\lambda k}{n^{1/3}} + \tfrac{\lambda c^2}{2} + o(1) \\ &= -\tfrac{c^3}{6} - \tfrac{\lambda^2 c}{2} + \tfrac{\lambda c^2}{2} + o(1) \end{aligned}$$

so that

$$E[X] \sim n^{-2/3} e^{-\frac{c^3}{6} - \frac{\lambda^2 c}{2} + \frac{\lambda c^2}{2}} c^{-5/2}(2\pi)^{-1/2}$$

For any particular such k $E[X] \to 0$ but if we sum k between $cn^{2/3}$ and $(c+dc)n^{2/3}$ we multiply by $n^{2/3}dc$. Going to the limit gives an integral: For any fixed a, b, λ let X be the number of tree components of size between $an^{2/3}$ and $bn^{2/3}$. Then

$$\lim_{n\to\infty} E[X] = \int_a^b e^{-\frac{c^3}{6} - \frac{\lambda^2 c}{2} + \frac{\lambda c^2}{2}} c^{-5/2}(2\pi)^{-1/2} dc$$

The large components are not all trees. E.M. Wright [1977] proved that for fixed l there are asymptotically $c_l k^{k-2+\frac{3}{2}l}$ connected graphs on k points for fixed k. For $c < 1$

$$E[Y] = \sum E[Y_k] \to \sum_{k=1}^{\infty} \frac{c^k}{2k}$$

has a finite limit whereas for $c > 1$, $E[Y] \to \infty$. Even for $c > 1$ for any fixed k the number of k-cycles has a limiting expectation and so do not asymptotically affect the number of components of a given size.

3 Inside the Phase Transition

In the evolution of the random graph $G(n,p)$ a crucial change takes place in the vicinity of $p = c/n$ with $c = 1$. The small components at that time are rapidly joining together to form a giant component. This corresponds to the Branching Process when births are Poisson with mean 1. There the number T of organisms will be finite almost always and yet have infinite expectation. No wonder that the situation for random graphs is extremely delicate. In recent years there has been much interest in looking "inside" the phase transition at the growth of the largest components. (See, e.g. Luczak [1990].) The appropriate parametrization is, perhaps surprisingly,

$$p = \frac{1}{n} + \frac{\lambda}{n^{4/3}}$$

When $\lambda = \lambda(n) \to -\infty$ the phase transition has not yet started. The largest components are $o(n^{2/3})$ and there are many components of nearly the largest size. When $\lambda = \lambda(n) \to +\infty$ the phase transition is over - a largest component, of size $>> n^{2/3}$ has emerged and all other components are of size $o(n^{2/3})$. Let's fix λ, c and let X be the number of tree components of size $k = cn^{2/3}$. Then

$$E[X] = \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)+\binom{k}{2}-(k-1)}$$

Watch the terms cancel!

$$\binom{n}{k} = \frac{(n)_k}{k!} \sim \frac{n^k e^k}{k^k \sqrt{2\pi k}} \prod_{i=1}^{k-1} (1 - \frac{i}{n})$$

For $i < k$

$$-\ln(1 - \frac{i}{n}) = \frac{i}{n} + \frac{i^2}{2n^2} + O(\frac{i^3}{n^3})$$

stage, as only small components have thus far been found, the number of remaining points is $m = n - O(1) \sim n$ so the conditional probabilities of small, giant and failure remain asymptotically the same. The chance of ever hitting a failure component is thus $\leq s\epsilon$ and the chance of hitting all small components is $\leq (y+\epsilon)^s \leq \epsilon$ so that with probability at least $1-\epsilon'$, where $\epsilon' = (s+1)\epsilon$ may be made arbitrarily small, we find a series of less than s small components followed by a giant component. The remaining graph has $m \sim yn$ points and $pm \sim cy = d$, the conjugate of c as defined earlier. As $d < 1$ the previous analysis gives the maximal components. In summary: almost always $G(n, c/n)$ has a giant component of size $\sim (1-y)n$ and all other components of size $O(\ln n)$. Furthermore, the Duality Principle has a discrete analog.

Discrete Duality Principle. Let $d < 1 < c$ be conjugates. The structure of $G(n, c/n)$ with its giant component removed is basically that of $G(m, d/m)$ where m, the number of vertices not in the giant component, satisfies $m \sim ny$.

The small components of $G(n, c/n)$ can also be examined from a static view. For a fixed k let X be the number of tree components of size k. Then

$$E[X] = \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)+\binom{k}{2}-(k-1)}$$

Here we use the nontrivial fact, due to Cayley, that there are k^{k-2} possible trees on a given k-set. For c, k fixed

$$E[X] \sim n \frac{e^{-ck} k^{k-2} c^{k-1}}{k!}$$

As trees are strictly balanced a second moment method gives $X \sim E[X]$ almost always. Thus $\sim p_k n$ points lie in tree components of size k where

$$p_k = \frac{e^{-ck}(ck)^{k-1}}{k!}$$

It can be shown analytically that $p_k = \Pr[T = k]$ in the Branching Process with mean c. Let Y_k denote the number of cycles of size k and Y the total number of cycles. Then

$$E[Y_k] = \frac{(n)_k}{2k} (\frac{c}{n})^k \sim \frac{c^k}{2k}$$

There are n choices for initial vertex v. Thus almost always *all* components have size $O(\ln n)$.

Now assume $c > 1$. For any fixed t, $\lim_{n\to\infty} \Pr[T = t] = \Pr[T^* = t]$ but what corresponds to $T^* = \infty$? For $t = o(n)$ we may estimate $1-(1-p)^t \sim pt$ and $n-1 \sim n$ so that

$$\Pr[Y_t \leq 0] = \Pr[B[n-1, 1-(1-p)^t] \leq t-1] \sim \Pr[B[n, tc/n] \leq t]$$

drops exponentially in t by Large Deviation results. When $t = \alpha n$ we estimate $1-(1-p)^t$ by $1-e^{-c\alpha}$. The equation $1-e^{-c\alpha} = \alpha$ has solution $\alpha = 1-y$ where y is the extinction probability. For $\alpha < 1-y$, $1-e^{-c\alpha} > \alpha$ and

$$\Pr[Y_t \leq 0] \sim \Pr[B[n, 1-e^{-c\alpha}] \leq \alpha n]$$

is exponentially small while for $\alpha > 1-y$, $1-e^{-c\alpha} < \alpha$ and $\Pr[Y_t \leq 0] \sim 1$. Thus almost always $Y_t = 0$ for some $t \sim (1-y)n$. Basically, $T^* = \infty$ corresponds to $T \sim (1-y)n$. Let $\epsilon, \delta > 0$ be arbitrarily small. With somewhat more care to the bounds we may show that there exists t_0 so that for n sufficiently large

$$\Pr[t_0 < T < (1-\delta)n(1-y) \;\; \text{or} \;\; \mathrm{T} > (1+\delta)\mathrm{n}(1-\mathrm{y})] < \epsilon$$

Pick t_0 sufficiently large so that

$$y - \epsilon \leq \Pr[T^* \leq t_0] \leq y$$

Then as $\lim_{n\to\infty} \Pr[T \leq t_0] = \Pr[T^* \leq 0]$ for n sufficiently large

$$y - 2\epsilon \leq \Pr[T \leq t_0] \leq y + \epsilon$$

$$1 - y - 2\epsilon \leq \Pr[(1-\delta)n(1-y) < T < (1+\delta)n(1-y)] < 1 - y + 3\epsilon$$

Now we expand our procedure to find graph components. We start with $G \sim G(n,p)$, select $v = v_1 \in G$ and compute $C(v_1)$ as before. Then we delete $C(v_1)$, pick $v_2 \in G - C(v_1)$ and iterate. At each stage the remaining graph has distribution $G(m,p)$ where m is the number of vertices. (Note, critically, that no pairs $\{w, w'\}$ in the remaining graph have been examined and so it retains its distribution.) Call a component $C(v)$ small if $|C(v)| \leq t_0$, giant if $(1-\delta)n(1-y) < |C(v)| < (1+\delta)n(1-y)$ and otherwise failure. Pick $s = s(\epsilon)$ with $(y+\epsilon)^s < \epsilon$. (For ϵ small $s \sim K \ln \epsilon^{-1}$.) Begin this procedure with the full graph and terminate it when either a giant component or a failure component is found or when s small components are found. At each the result follows by induction. □

We set $p = c/n$. When t and Y_{t-1} are small we may approximate Z_t by $B[n, c/n]$ which is approximately Poisson with mean c. Basically small components will have size distribution as in the Branching Process. The analogy must break down for $c > 1$ as the Branching Process may have an infinite population whereas $|C(v)|$ is surely at most n. Essentially, those v for which the Branching Process for $C(v)$ does not "die early" all join together to form the giant component.

Fix c. Let $Y_0^*, Y_1^*, \ldots, T^*, Z_1^*, Z_2^*, \ldots, H^*$ refer to the Branching Process and $Y_0, Y_1, \ldots, T, Z_1, Z_2, \ldots, H$ refer to the Random Graph process. For any possible history $(z_1, \ldots, z_t)$

$$\Pr[H^* = (z_1, \ldots, z_t)] = \prod_{i=1}^{t} \Pr[Z^* = z_i]$$

where Z^* is Poisson with mean c while

$$\Pr[H = (z_1, \ldots, z_t)] = \prod_{i=1}^{t} \Pr[Z_i = z_i]$$

where Z_i has Binomial Distribution $B[n-1-z_1-\ldots-z_{i-1}, c/n]$. The Poisson distribution is the limiting distribution of Binomials. When $m = m(n) \sim n$ and c, i are fixed

$$\lim_{n\to\infty} \Pr[B[m, c/n] = i] = \lim_{n\to\infty} \binom{m}{z} (\frac{c}{n})^z (1 - \frac{c}{n})^{m-z} = e^{-c} c^z / z!$$

hence

$$\lim_{n\to\infty} \Pr[H = (z_1, \ldots, z_t)] = \Pr[H^* = (z_1, \ldots, z_t)]$$

Assume $c < 1$. For any fixed t, $\lim_{n\to\infty} \Pr[T = t] = \Pr[T^* = t]$. We now bound the size of the largest component. For any t

$$\Pr[T > t] \le \Pr[Y_t > 0] = \Pr[B[n-1, 1-(1-p)^t] \ge t] \le \Pr[B[n, tc/n] \ge t]$$

as $1 - (1-p)^t \le tp$ and $n - 1 < n$. By Large Deviation Results

$$\Pr[T > t] < e^{-\alpha t}$$

where $\alpha = \alpha(c) > 0$. Let $\beta = \beta(c)$ satisfy $\alpha\beta > 1$. Then

$$\Pr[T > \beta \ln n] < n^{-\alpha\beta} = o(n^{-1})$$

2 The Giant Component

Now let's return to random graphs. We define a procedure to find the component $C(v)$ containing a given vertex v in a given graph G. We are motivated by Karp [1990] in which this approach is applied to random digraphs. In this procedure vertices will be live, dead or neutral. Originally v is live and all other vertices are neutral, time $t = 0$ and $Y_0 = 1$. Each time unit t we take a live vertex w and check all pairs $\{w, w'\}$, w' neutral, for membership in G. If $\{w, w'\} \in G$ we make w' live, otherwise it stays neutral. After searching all neutral w' we set w dead and let Y_t equal the new number of live vertices. When there are no live vertices the process terminates and $C(v)$ is the set of dead vertices. Let Z_t be the number of w' with $\{w, w'\} \in G$ so that

$$Y_0 = 1$$

$$Y_t = Y_{t-1} + Z_t - 1$$

With $G = G(n, p)$ each neutral w' has independent probability p of becoming live. Here, critically, no pair $\{w, w'\}$ is ever examined twice so that the conditional probability for $\{w, w'\} \in G$ is always p. As $t - 1$ vertices are dead and Y_{t-1} are live

$$Z_t \sim B[n - (t-1) - Y_{t-1}, p]$$

Let T be the least t for which $Y_t = 0$. Then $T = |C(v)|$. As in Section 1 we continue the recursive definition of Y_t, this time for $0 \leq t \leq n$.
Claim 2.1 For all t

$$Y_t \sim B[n-1, 1-(1-p)^t] + 1 - t$$

It is more convenient to deal with

$$N_t = n - t - Y_t$$

the number of neutral vertices at time t and show, equivalently,

$$N_t \sim B[n-1, (1-p)^t]$$

This is reasonable since each $w \neq v$ has independent probability $(1-p)^t$ of staying neutral t times. Formally, as $N_0 = n - 1$ and

$$\begin{aligned} N_t &= n - t - Y_t &&= n - t - B[n-(t-1)-Y_{t-1}, p] - Y_{t-1} + 1 \\ & &&= N_{t-1} - B[N_{t-1}, p] \\ & &&= B[N_{t-1}, 1-p] \end{aligned}$$

been defined. T (whether finite or infinite) is the total number of organisms, including the original, in this process. (A natural approach, found in many probability texts, is to have all organisms of a given generation have their children at once and study the number of children of each generation. While we may think of the organisms giving birth by generation it will not affect our model.)

We shall use the major result of Branching Processes that when $E[Z] = c < 1$ with probability one the process dies out ($T < \infty$) but when $E[Z] = c > 1$ then there is a nonzero probability that the process goes on forever ($T = \infty$).

When a branching process dies we call $H = (Z_1, \ldots, Z_T)$ the *history* of the process. A sequence $(z_1, \ldots, z_t)$ is a possible history if and only if the sequence y_i given by $y_0 = 1, y_i = y_{i-1} + z_i - 1$ has $y_i > 0$ for $0 \leq i < t$ and $y_t = 0$. When Z is Poisson with mean λ

$$\Pr[H = (z_1, \ldots, z_t)] = \prod_{i-1}^{t} \frac{e^{-\lambda}\lambda^{z_i}}{z_i!} = \frac{e^{-\lambda}(\lambda e^{-\lambda})^{t-1}}{\prod_{i=1}^{t} z_i!}$$

since $z_1 + \ldots + z_t = t - 1$.

We call $d < 1 < c$ a conjugate pair if

$$de^{-d} = ce^{-c}$$

The function $f(x) = xe^{-x}$ increases from 0 to e^{-1} in [0,1) and decreases back to 0 in $(1, \infty)$ so that all $c \neq 1$ have a uniqe conjugate. Let $c > 1$ and $y = \Pr[T < \infty]$ so that $y = e^{c(y-1)}$. Then $(cy)e^{-cy} = ce^{-c}$ so

$$d = cy$$

Duality Principle. Let $d < 1 < c$ be conjugates. The Branching Process with mean c, conditional on extinction, has the same distribution as the Branching Process with mean d.

Proof. It suffices to show that for every history $H = (z_1, \ldots, z_t)$

$$\frac{e^{-c}(ce^{-c})^{t-1}}{y \prod_{i=1}^{t} z_i!} = \frac{e^{-d}(de^{-d})^{t-1}}{\prod_{i=1}^{t} z_i!}$$

This is immediate as $ce^{-c} = de^{-d}$ and $ye^{-d} = ye^{-cy} = e^{-c}$.

Lecture 7: The Phase Transition

1 Branching Processes

Paul Erdős and Alfred Rényi, in their original 1960 paper, discovered that the random graph $G(n,p)$ undergoes a remarkable change at $p = 1/n$. Speaking roughly, let first $p = c/n$ with $c < 1$. Then $G(n,p)$ will consist of small components, the largest of which is of size $\Theta(\ln n)$. But now suppose $p = c/n$ with $c > 1$. In that short amount of "time" many of the components will have joined together to form a "giant component" of size $\Theta(n)$. The remaining vertices are still in small components, the largest of which has size $\Theta(\ln n)$. They dubbed this phenomenon the *Double Jump*. We prefer the descriptive term Phase Transition because of the connections to percolation (e.g., freezing) in mathematical physics.

To better understand the Phase Transition we make a lengthy detour into the subject of Branching Processes. Imagine that we are in a unisexual universe and we start with a single organism. Imagine that this organism has a number of children given by a given random variable Z. (For us, Z will be Poisson with mean c.) These children then themselves have children, the number again being determined by Z. These grandchildren then have children, etc. As $Z = 0$ will have nonzero probability there will be some chance that the line dies out entirely. We want to study the total number of organisms in this process, with particular eye to whether or not the process continues forever. (The original application of this model was to a study of the -gasp!- male line of British peerage.)

Now lets be more precise. Let $Z_1, Z_2, \ldots$ be independent random variables, each with distribution Z. Define $Y_0, Y_1, \ldots$ by the recursion

$$Y_0 = 1$$

$$Y_i = Y_{i-1} + Z_i - 1$$

and let T be the least t for which $Y_t = 0$. If no such t exists (the line continuing forever) we say $T = +\infty$. The Y_i and Z_i mirror the Branching Process as follows. We view all organisms as living or dead. Initially there is one live organism and no dead ones. At each time unit we select one of the live organisms, it has Z_i children, and then it dies. The number Y_i of live organisms at time i is then given by the recursion. The process stops when $Y_t = 0$ (extinction) but it is a convenient fiction to define the recursion for all t. Note that T is not affected by this fiction since once $Y_t = 0$, T has