3.1.6 The emerging graph

Consider G(n,p) as p grows. Starting with $p=0$, i.e., a graph of n vertices and no edges, as p increases and edges are added, a forest of trees emerges. The graph is a forest of trees for $p \in o(1/n)$; we showed this already by proving that for such p, almost surely, there are no cycles. When p becomes asymptotically equal to $\frac{1}{n}$, cycles begin to appear but each component is either a tree or is unicyclic, (i.e. the component contains at most one cycle) with no component having more than $\log n$ vertices. The number of components containing a single cycle is a constant independent of n. Thus, the graph consists of a forest of trees plus a few components that have a single cycle. No component has more than $\log n$ vertices

As p approaches $\frac{1}{n}$, a phase transition occurs in which a giant component emerges. The transition consists of a double jump. At $p=\frac{1}{n}$ a giant components of $n^{\frac{2}{3}}$ vertices emerge, which are almost surely trees. Then at $p=\frac{d}{n}$, d>1, a true giant component emerges that has a number of vertices proportional to n. As p increases further, all non isolated vertices are absorbed into the giant component and at $p=\frac{1}{4}\frac{\ln n}{n}$ the graph consists only of isolated vertices plus a giant component. At $p=\frac{\ln n}{n}$ the graph becomes completely connected. By $p=\frac{1}{2}$ the graph is not only connected but is sufficiently dense so that it has a clique of size $(2-\varepsilon)\log n$ for any $\varepsilon>0$.

The region $p=\frac{d}{n}$, d≤1.

Let x_m be the number of components of G that are trees with at least m vertices. Now

$$E(x_m)=\sum_{k=m}^{n}\binom{n}{k}k^{k-2}p^{k-1}(1-p)^{\binom{k}{2}-(k-1)+k(n-k)}.$$

[To specify a tree, first select k out of the n vertices. Given the k vertices, there are k^{k-2} possible ways to form a tree. This is a basic fact which we do not prove here. Once a tree is formed each of $k-1$ edges of the tree is present with probability p. Each of the other $\binom{k}{2}-(k-1)$ edges among the k vertices of the tree is absent with probability $1-p$ and each of the $k(n-k)$ edges from the k vertices of the tree to the remaining $n-k$ vertices is absent with probability $1-p$.]

We now make a series of approximations to derive a formula for $E(x_m)$ assuming again $p=\frac{d}{n}$ with $d\le 1$.

$$(1-p)^{\binom{k}{2}-(k-1)+k(n-k)}\cong\theta(1)\left[\left(1-\tfrac{d}{n}\right)^{n-\frac{k}{2}}\right]^k\cong\theta(1)\left[e^{\frac{kd}{2n}}e^{-d}\right]^k$$

Since $k\le n$ the term $\left(1-\frac{d}{n}\right)^{-\frac{3}{2}k+1}$ is bounded above and below by a constant and it was replaced in the above equation by the $\theta(1)$ term.

Next since $k! \cong \frac{k^{k+\frac{1}{2}}}{e^k}$, write $\binom{n}{k} = \frac{n(n-1)\cdots(n-k)}{k!} \cong \frac{n(n-1)\cdots(n-k)}{k^{k+\frac{1}{2}}} e^k$. Substituting into $E(x_m)$ (and using the abbreviation $\cong$ here to mean within a constant factor), we get

$$E(x_m) \cong \sum_{k=m}^{n} \frac{n(n-1)\cdots(n-k)}{k^{k+\frac{1}{2}}} e^k k^{k-2} \left(\frac{d}{n}\right)^{k-1} \left(\frac{e^{\frac{kd}{2n}}}{e^d}\right)^k$$

$$\cong \tfrac{n}{d} \sum_{k=m}^{n} \frac{n(n-1)\cdots(n-k)}{n^k} \frac{1}{k^{\frac{5}{2}}} \left[e^{\frac{dk}{2n}} d e^{1-d}\right]^k$$

Now $\frac{n(n-1)\cdots(n-k)}{n^k} e^{\frac{dk^2}{2n}} = O(1)$. [We see this by using $1 - \frac{t}{n} \le e^{-t/n}$ from which it follows that that the expression is at most $\exp(-\frac{k^2}{2n} + \frac{dk^2}{2n})$ which for $d \le 1$ is O(1).] Thus,

$$E(x_m) \cong \frac{n}{k^{\frac{5}{2}}} \left(de^{1-d}\right)^k \qquad \text{Eq. 3.1}$$

Theorem: In G(n,p) with $p = \frac{d}{n}$, d<1, almost surely, the tree components are of size $O(\log n)$.

Proof: From above $E(x_m) \cong \frac{n}{k^{\frac{5}{2}}} \left(de^{1-d}\right)^k$. Now for d<1, it follows that $0 < de^{1-d} < 1$. [Let $f(d) = de^{1-d}$. f(0)=0, f(1)=1, and $f' \ne 0$ for 0<d<1.] Let $c = de^{1-d}$. Since c<1, if *k* grows as $\log n$, $E(x_m) \cong \frac{c^{\log n}}{(\log n)^{\frac{5}{2}}} \to 0$. Thus, for $p < \frac{1}{n}$, all components that are trees have size $O(\log n)$.

■

For $p = \frac{1}{n}$ there is a phase transition in which a tree of size $n^{\frac{2}{3}}$ appears.

Theorem: For $p = \frac{1}{n}$, the maximum size of a tree component is $O(n^{\frac{2}{3}})$. Also, for some constant c>0, the expected number of tree components of size at least $cn^{\frac{2}{3}}$ tends to a positive real.

Proof: For $p = \frac{1}{n}$, from Eq. 3.1, we see that the expected number of trees with at least *m* vertices is $\theta(1) n \sum_{k=m}^{n} \frac{1}{k^{\frac{5}{2}}}$. The summation, $\sum_{k=m}^{n} \frac{1}{k^{\frac{5}{2}}}$, is very close to $\int_{k=m}^{n} k^{-5/2} dk$ which is

$(2/3)(m^{-3/2}-n^{-3/2})$. If $m > cn^{\frac{2}{3}}$, the summation converges to zero. Thus, with high probability no component is larger than $O(n^{\frac{2}{3}})$ proving the first statement. The second statement follows from the same calculation.

■

We have shown that for tree components the maximum size component for $p < \frac{1}{n}$ is $\log n$ and for $p = \frac{1}{n}$ is order $n^{\frac{2}{3}}$. When $p \le \frac{1}{n}$, almost all vertices are in trees and hence the maximum size tree is actually the maximum size component. We do not prove this here.

References

Material from "Graphical Evolution," Edgar M. Palmer

Appendix Material

Here we prove that the number of trees with n vertices is n^{n-2}. By a labeled tree we mean a tree with n vertices and n distinct labels, each label assigned to one vertex.

Theorem: The number of labeled trees with n vertices is n^{n-2}.

Proof: (Prüfer sequence) There is a one-to-one correspondence between labeled trees and sequences of length n-2 of integers between 1 and n. An integer may repeat in the sequence. The number of such sequences is clearly n^{n-2}. Although each vertex of the tree has a unique integer label the corresponding sequence has repeating labels. The reason for this is that the labels in the sequence refer to interior vertices of the tree and the number of times the integer corresponding to an interior vertex occurs in the sequence is related to the degree of the vertex. Integers corresponding to leaves do not appear in the sequence.

To see the one-to-one correspondence, first convert a tree to a sequence by deleting the lowest numbered leaf. If the lowest numbered leaf is i and its parent is j, append j to the tail of the sequence. Repeating the process until only two vertices remain yields the sequence. Clearly a labeled tree gives rise to only one sequence.

It remains to show how to construct a unique tree from a sequence. The proof is by induction on n. For n=1 or 2 the induction hypothesis is trivially true. Assume the induction hypothesis true for n-1, $n \geq 3$. Certain numbers from 1 to n do not appear in the sequence and these numbers correspond to vertices that are leaves. Let i be the lowest number not appearing in the sequence and let j be the first integer in the sequence. Then i corresponds to a leaf connected to vertex j. Delete the integer j from the sequence. Then by the induction hypothesis there is a unique labeled tree with integer labels $1,2,\cdots,i-1,i+1,\cdots,n$. Add the leaf i by connecting the leaf to vertex j. We need to argue that no other sequence can give rise to the same tree. Suppose some other sequence did. Then the i^{th} integer in the sequence must be j. By the induction hypothesis the sequence with j removed is unique. ■

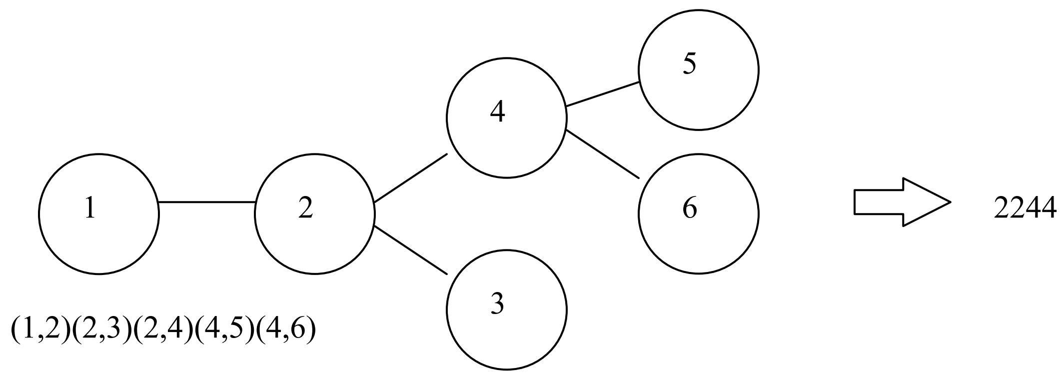
Algorithm

Create leaf list – the list of labels not appearing in the Prüfer sequence. n is the length of the Prüfer list plus two.

```
while Prüfer sequence is non empty do
begin
    p=first integer in Prüfer sequence
    e=smallest label in leaf list
    Add edge (p,e)
    Delete e from leaf list
    Delete p from Prüfer sequence
```

If p no longer appears in Prüfer sequence
add p to leaf list
end
there are two vertices e and f on leaf list, add edge (e,f)
finish fixing above program
Reference – Prüfer Wikipedia
http://www.cs.mcgill.ca/~abatko/computers/free_trees/project/

Example:



Exercise 3.35: Draw a tree with 10 vertices and label each vertex with a unique integer from 1 to 10. Construct the Prüfer sequence for the tree. Given the Prüfer sequence recreate the tree.

Exercise 3.36: Construct the tree corresponding to the following Prüfer sequences
a) 113663 (1,2),(1,3),(1,4),(3,5),(3,6),(6,7), and (6,8)
b) 552833226

See also Kirchkoff's theorem in Wikipedia for interesting material. Reachable from Purfer page.

Exercises

Exercise 3.37: Find a data base in machine readable form that can be viewed as a graph. Find the number of components of various sizes. Calculate *p*. Examine the small components and see if any have cycles.

References

Bollobas "Random Graphs"
This book is very hard to read, look at only if one needs a proof of a result.

Reference Janson, Luczak and Rucinski – easy to read with good material.

The birth of the giant component, Svante Janson, Donald E. Knuth, Tomasz Łuczak, Boris Pittel.
http://arxiv.org/PS_cache/math/pdf/9310/9310236v1.pdf