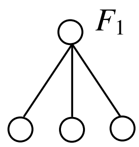
CS 485 Lecture 15 — 02/24/06

Branching Processes

We know that the process either goes to extinction or grows forever. However, it cannot be the case that the process comes to a stage where the number of children in the j^{th} generation hovers about a constant, c.

Random Walks

Examining the frontiers of a tree (i.e. by doing a breadth first search). Pick a vertex and let it be our first frontier, F_1.



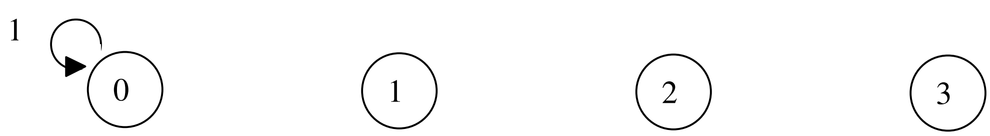
Explore F_1's children. Generalizing, we get:

$$F_{i+1} = F_i - 1 + X_i$$

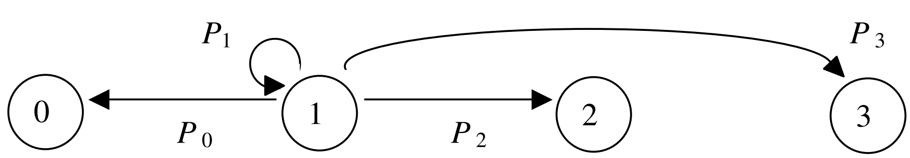
Suppose we represent the frontiers' sizes as follows:

Frontier size of 0 | Frontier size of 1 | Frontier size of 2 | Frontier size of 3

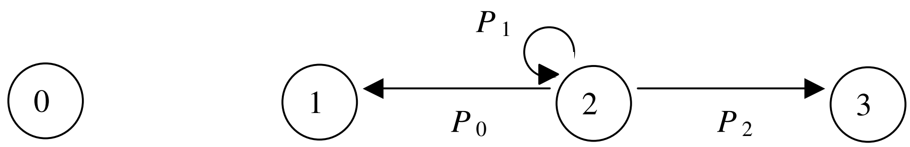
Then if our current frontier F_i is of size 0, we know with probability 1 that future frontiers will also have size 0.



Likewise, if our current frontier F_i is of size 1, then the probability that by exploring the children of F_i, the size of F_{i+1} is 1, 2, 3, etc is:



where P_0 is the probability of having 0 children, etc. And for a frontier F_i of size 2, we have:



If this random walk has a limiting/stationary distribution, what can it look like? The probability of being in vertex 1 cannot be nonzero, since if it were, this would change the probability of being in vertex 0 at the next time step, contradicting the fact that we have a stationary distribution. Likewise, No other vertex *c* for some constant *c* can have a nonzero probability. This leaves us with vertex 0 having some probability *q*, which need not be 1, just as we had deduced before.

<u>Sum of *n* Random Variables with Identical Distributions</u>

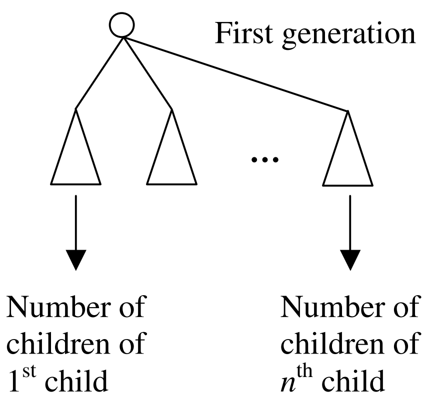
The expected value of the sum of *n* random variables, each with identical distributions is given by:

$$x_1 + x_2 + x_3 + \ldots + x_n \;= \mathrm{E}(x_1) + \mathrm{E}(x_2) + \mathrm{E}(x_3) + \ldots + \mathrm{E}(x_n)$$

and since their distributions are the same, the expected values of each individual x_i is also identical, giving:

$$x_1 + x_2 + x_3 + \ldots + x_n \;= \mathrm{n}(x_i)$$

Note that the above did not constrain the x_i's to be independent. What if n itself is a random variable? For example, what is the number of children in the 2nd generation below?



Where we know the expected number of children in the first generation is given by *n*. Furthermore, each of the *n* children has x_i children, where the x_i's have the same distribution as *n*.

Claim : Provided that *n* and the x_i's are independent of each other, we can write the expected number of children in the 2nd generation as:

$$E\left(\sum_{i=1}^{n} x_i\right) \qquad = E(n)E(x_1)$$

Proof :

$$E\left(\sum_{i=1}^{n} x_i\right) = \Pr[n = 1 \,\&\, x_1 = 1] + 2\Pr[n = 1 \,\&\, x_1 = 2] + \ldots + i\Pr[n = 1 \,\&\, x_1 = i] + \ldots$$
$$+ \Pr[n = 2 \,\&\, x_1 = 1] + 2\Pr[n = 2 \,\&\, x_1 = 2] + \ldots + i\Pr[n = 2 \,\&\, x_1 = i] + \ldots$$
$$+ \Pr[n = 2 \,\&\, x_2 = 1] + 2\Pr[n = 2 \,\&\, x_2 = 2] + \ldots + i\Pr[n = 2 \,\&\, x_2 = i] + \ldots$$

Since n and the x_i's are independent, $\Pr[n = a \,\&\, x_i = b]$ is equal to $\Pr[n = a]\Pr[x_i = b]$. Factoring out the $\Pr[n = a]$'s, we have:

$$E\left(\sum_{i=1}^{n} x_i\right) = \Pr[n = 1]\,(\Pr[x_1 = 1] + 2\Pr[x_1 = 2] + \ldots + i\Pr[x_1 = i] + \ldots) + \ldots$$
$$+ \Pr[n = 2]\,(\Pr[x_1 = 1] + 2\Pr[x_1 = 2] + \ldots + i\Pr[x_1 = i] + \ldots) + \ldots$$
$$+ \Pr[n = 2]\,(\Pr[x_2 = 1] + 2\Pr[x_2 = 2] + \ldots + i\Pr[x_2 = i] + \ldots) + \ldots$$

which is equivalent to

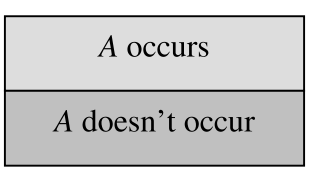
$$E\left(\sum_{i=1}^{n} x_i\right) = \Pr[n = 1]\,\mathrm{E}(x_1) + \Pr[n = 2]\,\mathrm{E}(x_1) + \Pr[n = 2]\,\mathrm{E}(x_2) + \ldots$$

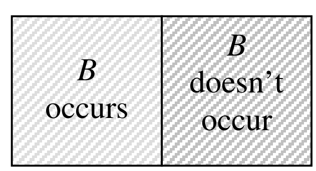
Finally, factoring out the $\mathrm{E}(x_i)$'s, we have:

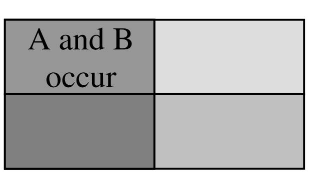
$$E\left(\sum_{i=1}^{n} x_i\right) = (\Pr[n = 1] + 2\Pr[n = 2] + 3\Pr[n = 3] + \ldots)\,\mathrm{E}(x_1)$$
$$= \mathrm{E}(n)\,\mathrm{E}(x_1)$$

<u>Note on Independence of Random Variables</u>

If we have 2 events A and B that are independent of each other and each occurs with probability p = 0.5, then A and B can be represented in the probability space as:







In this case, the probability that both *A* and *B* occur is simply the product of the probabilities that each occur. Pr[*A* and *B*] = Pr[*A*] Pr[*B*] = ¼ in our case.

This is not true if *A* and *B* are not independent. For example, if $A = \neg B$, then Pr[*A* and *B*] = 0.

<u>Building Graphs with Structure</u>

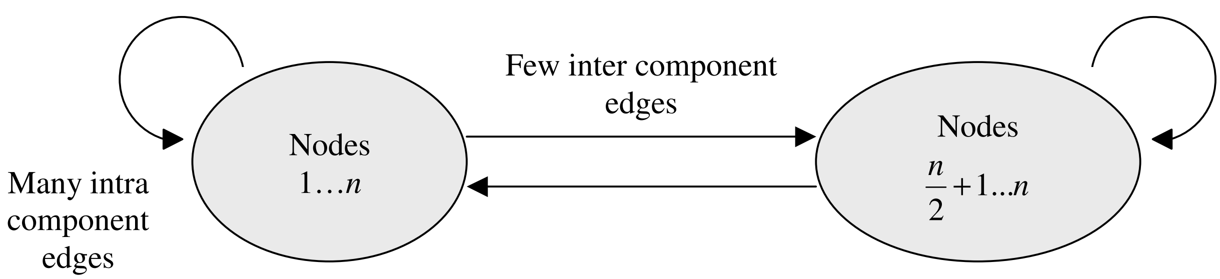
Consider the following *n* x *n* matrix, where each (*i*, *j*) cell is given value 1 with probability $\frac{1}{n}$ and 0 with probability $\frac{1-n}{n}$.

$$\begin{pmatrix} \\ \frac{1}{n} \\ \\ \end{pmatrix}$$

Suppose we now split this matrix into 4 smaller matrices, such that

$$\left(\begin{array}{c|c} \frac{1}{2} & \frac{1}{4} \\ \hline \frac{1}{4} & \frac{1}{2} \end{array}\right)$$

This results in a graph that has 2 components, where there are many edges between vertices within the same component and fewer edges between vertices from different components:



We could then such matrices with one another:

$$\begin{pmatrix} \\ \frac{1}{2} \\ \\ \end{pmatrix} \quad \text{with} \quad \left(\begin{array}{c|c} \frac{1}{2} & \frac{1}{4} \\ \hline \frac{1}{4} & \frac{1}{2} \end{array}\right)$$

Non-Uniform Degree Model

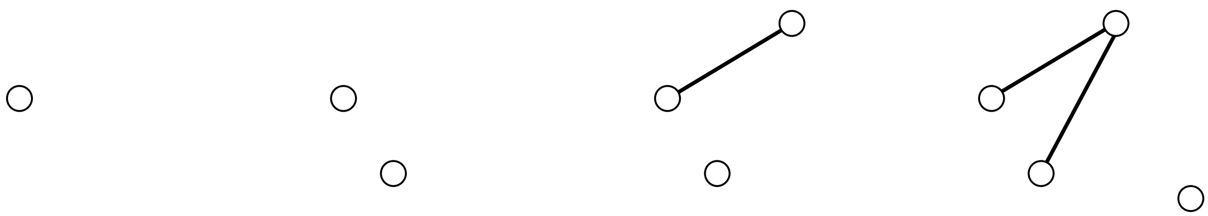
One way to create such a model is to fix the degree of each vertex:

$$\frac{\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \begin{pmatrix} & & \\ & \mathrm{M} & \\ & & \end{pmatrix} \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}}{\sum d_i}$$

where M is a structured matrix and d_i is the degree of vertex i.

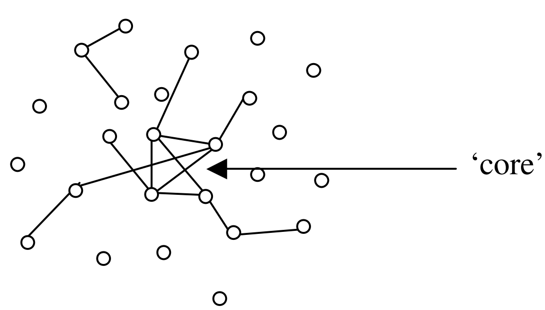
Grown Graph

Start with an empty graph. At each unit of time, add 1 vertex to the graph and pick two vertices u and v currently present in the graph and with probability p, add and edge between them.



$t = 1$ $t = 2$ $t = 3$ $t = 4$

One observation from this model is that there tends to be a core where the edge concentration is high. This is due to the fact that during the initial stages of the growing the graph, only a few vertices are present. Hence, the probability of edges being present within this few vertices are high, and therefore the density of edges will also be high.



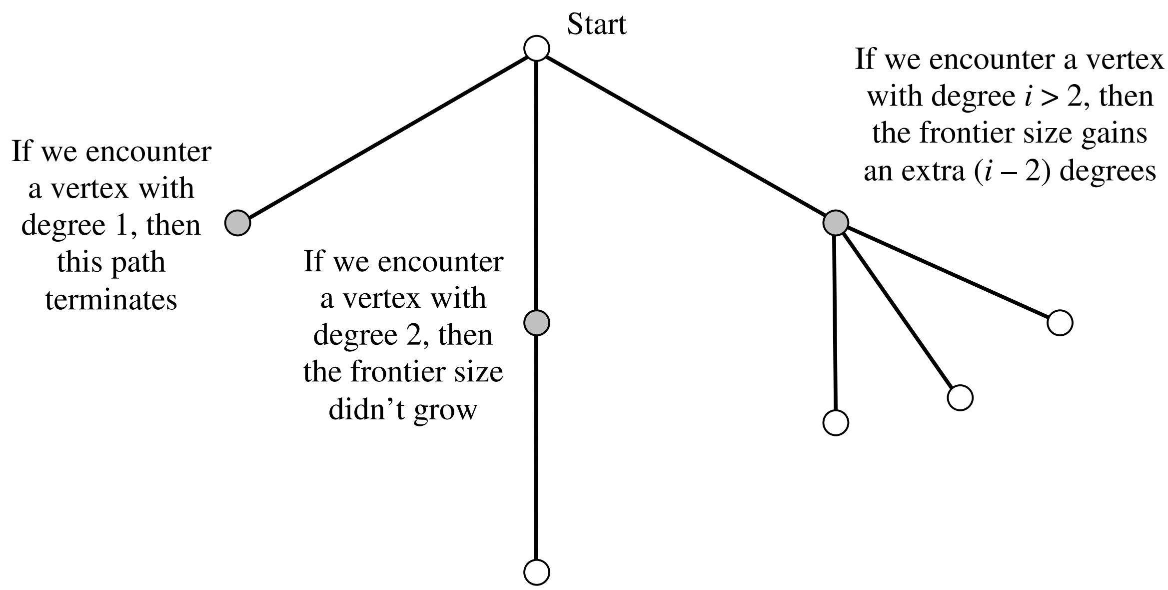
Random Graph with Given Degree Distribution

Let λ_i be the fraction of the vertices with degree *i*.

Claim : A phase transition occurs when

$$\sum_{i=1}^{n} i(i-2)\lambda_i \quad = 0$$

Suppose we start off at a vertex, and explore its children.



Therefore, the $(i - 2)$ part of the above equation represents the net gain in the frontier's size. The factor of λ_i makes sense since we have to account for how many vertices of a degree we have in our graph. However, why the remaining factor of *i*? Loosely, the frequency with which we will encounter a vertex of a certain degree depends not only on how many such vertices are in the graph, but on the degree itself. This is because there are more ways to get to a vertex with higher degree than it is to get to a vertex with a low degree.

Other Models

Other models include adding edges by selecting 2 vertices proportionate to their current degrees. This way, nodes with higher degrees go even higher ('rich gets richer').