Problem 1

Here we will show that the expected depth of a node n in a general 1-out preferential attachment graph is at most $c \log n$ for some constant $c > 0$.

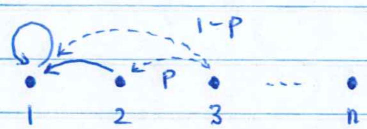
First recall that a general 1-out preferential attachment graph is defined by introducing nodes one at a time and assigned one edge from the node according to the selection rules

1. Set $dest(1) = 1$
2. For $t = 2 \ldots n$

   (a) Choose i uniformly in $\{1, \ldots, t-1\}$

   (b) Set $dest(i) = \begin{cases} i & \text{prob } p \\ dest(i) & \text{prob } 1-p \end{cases}$

which has the following picture



where the expected depth of a node n is the expected number of steps required to reach node 1. Note that the graph is a tree in structure and there can be no loops or cycles (other than 1).

We will do this in two ways, one which models stepping from node n to node 1 using a Dirichlet process and one which uses a recursive expression for depth and induction instead.

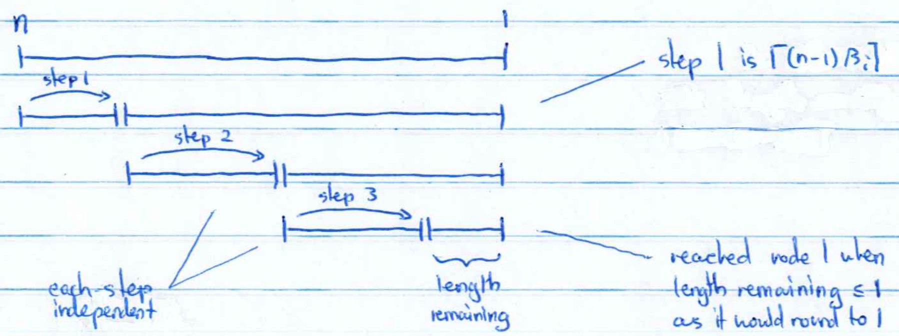
Dirichlet Process Method

First note that we only need an upper bound on the expected depth so we can consider the worst case scenario when $p = 1$.

Note that when $p=0$ the selection rules will always choose node 1 as $dest(1)=1$ and for $t=2 \ldots n$ we have $dest(t) = dest(i)$ and so recursively $dest(t) = 1$.

But when $p=1$ the selection rules will always choose a node and not the destination of that node which maximises the expected number of steps required to reach node 1.

Consider the following picture of the stepping model



where each step can be modelled using $Unif(0,1)$ which is scaled to the length remaining, which is known as the stick-breaking process and is one of the standard Dirichlet process constructions.

Hence if we define $\beta_i \sim Unif(0,1)$ as the proportional step size for step i starting from node n then the length remaining after k steps is at most the following

$$(n-1)\prod_{i=1}^{k}(1-\beta_i)$$

as each step would round down to the nearest integer node instead.

Hence the expected length remaining after k steps can be bounded

$$\mathbb{E}[\text{length remaining}] \leq (n-1)\prod_{i=1}^{k}\mathbb{E}(1-\beta_i)$$

as each step is independent as a proportion of the length remaining.

Evaluating this noting that B_i Unif$(0,1)$

$$\mathbb{E}[\text{length remaining}] \leq (n-1) \prod_{i=1}^{k} (1 - \mathbb{E}(B_i))$$

$$= (n-1) \prod_{i=1}^{k} \frac{1}{2}$$

$$= \frac{n-1}{2k}$$

and noting that when $\mathbb{E}[\text{length remaining}] < 1$ we have reached node 1 as the last step would round down to node 1 we have the following

$$\frac{n-1}{2^k} < 1$$

$$k > \log(n-1)$$

and so after $\log(n-1)$ steps we expect to have reached node 1.

From this and the definition of expected depth we have the bound

$$\mathbb{E}[\text{depth node } n] = \mathbb{E}[\text{steps to node } 1]$$

$$\leq 1 + \log(n-1)$$

$$\leq c \log(n)$$

where c satisfies the following bound

$$c \geq \frac{1}{\log(n)} + \frac{\log(n-1)}{\log(n)}$$

which has a maximum value of $c \geq 1.29248$ when $n = 4$ and converges to $c \geq 1$ as $n \to \infty$ as expected.

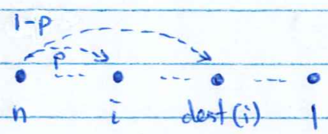
Hence the expected depth of node n in a general 1-out preferential attachment graph is at most $c \log n$ for some constant $c \geq 1.29248$.

However to tighten up the somewhat loose treatment of rounding and equivalent definitions of depth and steps and length remaining we can use a recursive expression for depth and induction instead.

Induction Method

Here we will consider the general case for any p as it will reduce to the $p = 1$ case when we take an upper bound in the recursive expression as desired.

Consider the following picture of one edge choice in the process



where we first choose i uniformly in $\{n-1, \ldots, 1\}$ and then choose the edge to be i with probability p or $dest(i)$ with probability $1-p$ as described.

Let X_i be the depth of node i for $i = n-1 \ldots 1$. Then the expected depth of node n or $\mathbb{E}[X_n]$ is related to the expected depth of all other nodes $i = n-1 \ldots 1$ by the following recursive expression

$$\begin{aligned}
\mathbb{E}[X_n] &= \sum_{i=1}^{n-1}\left[\frac{1}{n-1}\Big(p(\mathbb{E}[X_i]+1) + (1-p)(\mathbb{E}[X_{dest(i)}]+1)\Big)\right] \\
&= 1 + \frac{1}{n-1}\sum_{i=1}^{n-1}\Big(p\,\mathbb{E}[X_i] + (1-p)\,\mathbb{E}[X_{dest(i)}]\Big) \\
&\leq 1 + \frac{1}{n-1}\sum_{i=1}^{n-1}\mathbb{E}[X_i]
\end{aligned}$$

where $\mathbb{E}[X_i] \geq \mathbb{E}[X_{dest(i)}]$ is trivial.

From this we can find an explicit solution for $\mathbb{E}[X_n]$ which satisfies the above inequality with equality (or is just an exact solution for $\mathbb{E}[X_n]$ when $p = 1$ which is the worst case scenario still).

Suppose the ansatz for $\mathbb{E}[X_n]$

$$\mathbb{E}[X_n] = \sum_{k=1}^{n-1} \frac{1}{k}$$

which holds for all $n \geq 2$ where $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 1$ is trivial.

Now assume the ansatz holds up to node $n-1$ and assume $p = 1$

$$\mathbb{E}[X_n] = 1 + \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{E}[X_i]$$

$$= 1 + \frac{1}{n-1} \sum_{i=1}^{n-1} \sum_{k=1}^{i-1} \frac{1}{k}$$

$$= 1 + \frac{1}{n-1} \left[\frac{n-2}{1} + \frac{n-3}{2} + \dots + \frac{1}{n-2} \right]$$

$$= 1 + \frac{1}{n-1} \left[n \sum_{k=1}^{n-2} \frac{1}{k} - (n-2) - \sum_{k=1}^{n-2} \frac{1}{k} \right]$$

$$= 1 + \sum_{k=1}^{n-2} \frac{1}{k} - \frac{n-2}{n-1}$$

$$= \sum_{k=1}^{n-2} \frac{1}{k} + \frac{1}{n-1}$$

$$= \sum_{k=1}^{n-1} \frac{1}{k}$$

and so the ansatz holds for node n as well and therefore for all $i = n-1 \dots 1$ and for all n as well.

Hence we only need to bound the partial sum of the well known harmonic series and we can use one of the standard bounds below

$$\sum_{k=1}^{n-1} \frac{1}{k} \leq \gamma + \log(n-1)$$

$$\leq 1 + \log(n-1)$$

where $\gamma = 0.577215$ is Euler's constant and where the upper bound involving 1 can be proved using an integral bound which is trivial.

Hence we have the same bound

$$\mathbb{E}[X_n] \leq 1 + \log(n-1)$$

10/10

which is made only slightly tighter using Euler's constant instead

$$\mathbb{E}[X_n] \leq \gamma + \log(n-1)$$

$$\leq c\log(n)$$

where c satisfies the bound

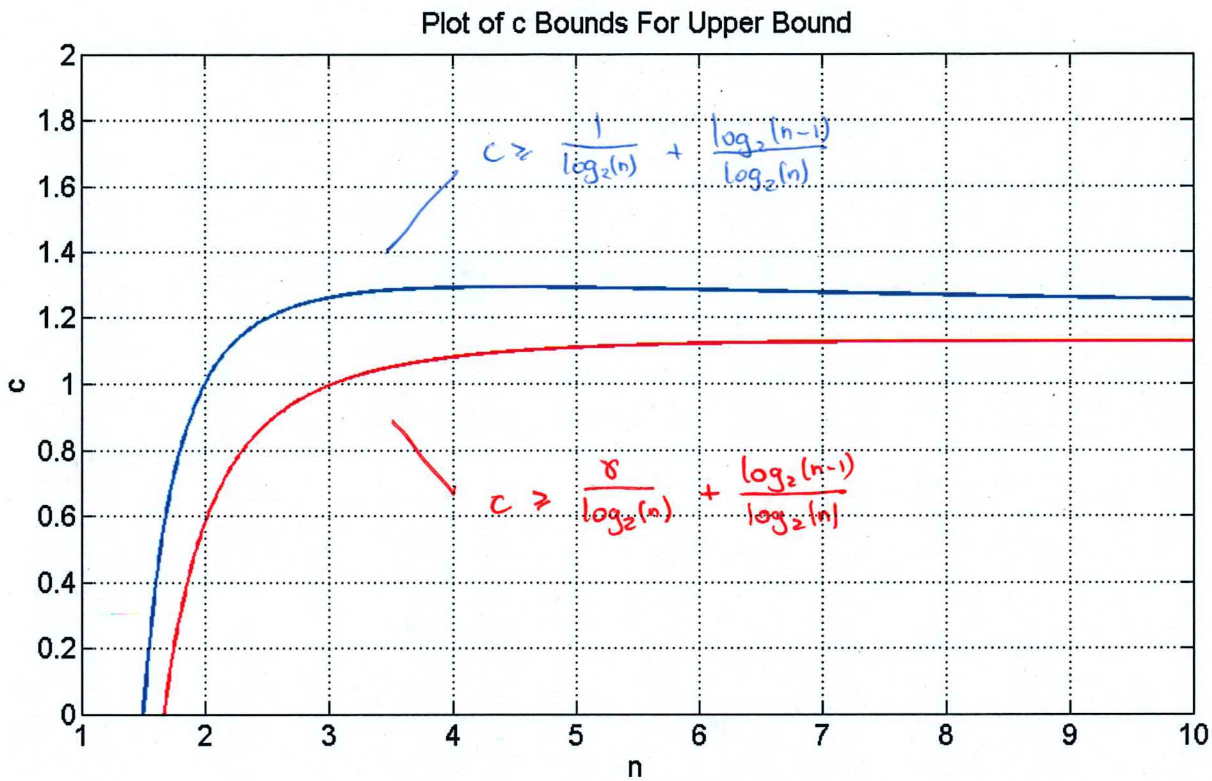
$$c \geq \frac{\gamma}{\log(n)} + \frac{\log(n-1)}{\log(n)}$$

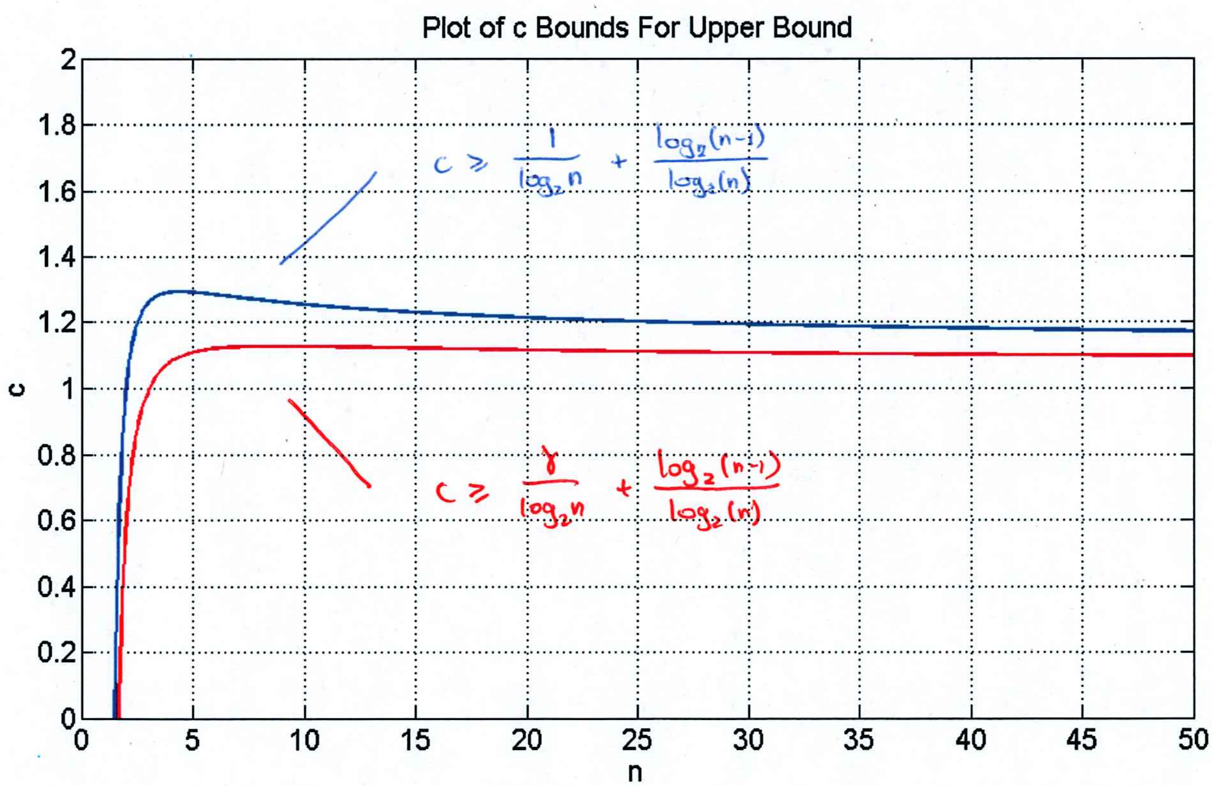
which has a maximum value of $c \geq 1.12848$ when $n=9$ and conveges to $c \geq 1$ as $n \to \infty$ as expected.

Hence the expected depth of node n in a general 1-out preferential attachment graph is at most $c \log n$ for some constant $c \geq 1.12848$.

Note that either of these methods is sufficient and as the bound is only the worst upper bound the size of c is not important as long as it is constant!

Also, the Dirichlet process method when $p=1$ only used Unif(0,1) to model the step where in general Beta$(1,\alpha)$ can be used where α determines the shape of the distribution and $\alpha < 1$ will bias the steps to being longer than the uniform Beta(1,1) steps which might allow us to finder an asymptotically tighter bound when $p < 1$ as well.


Plot of c Bounds For Upper Bound
c
n
$c \geq \frac{1}{\log_2(n)} + \frac{\log_2(n-1)}{\log_2(n)}$
$c \geq \frac{\gamma}{\log_2(n)} + \frac{\log_2(n-1)}{\log_2(n)}$


Plot of c Bounds For Upper Bound
c
n
$c \geq \frac{1}{\log_2 n} + \frac{\log_2(n-1)}{\log_2(n)}$
$c \geq \frac{\gamma}{\log_2 n} + \frac{\log_2(n-1)}{\log_2(n)}$