# Social Computing - Ex 1

Jonathan Hirsch - 200357697

May 20, 2015

# 1

## 1.1

Let $p < \frac{1}{4}, c > 0$, and examine the left subtree (meaning, the $\frac{n}{2}$ leaves on the left side of the tree. For each of these leaves, since the lowest common ancestor is the root then the probability to not be connected to any leaf on the right side is:

$$P_{not}\left(l,r\right) = \left(1 - c \cdot p^h\right)^{\frac{n}{2}}$$

and the probability for all leaves on the left to be disconnected from the right is

$$\begin{aligned}
P\left(L,R\right) &= \left[\left(1 - c \cdot p^h\right)^{\frac{n}{2}}\right]^{\frac{n}{2}} \\
&= \left(1 - c \cdot p^h\right)^{\frac{n^2}{4}} \\
\left(\text{plugging in } n = 2^h\right) &= \left(1 - c \cdot p^h\right)^{4^{h-1}} \\
&\approx e^{-c \cdot p^h 4^{h-1}} \\
&= e^{-\frac{c}{4}(4p)^h} \\
\left(p < \frac{1}{4} \Rightarrow 4p < 1 \Rightarrow\right) \lim_{h\to\infty} e^{-\frac{c}{4}(4p)^h} &\to e^0 \\
&= 1
\end{aligned}$$

### 1.2

Let $p = \frac{1}{2}, c > 0$. For each leaf, the number of possible neighbours at tree distance $d$ is $2^{d-1}$, then

$$\begin{aligned}
\mathbb{E}\left[\#_{neighbours}\right] &= \sum_{d=1}^{d=h} 2^{d-1} \cdot c \cdot p^d \\
&= c \sum_{d=1}^{d=h} 2^{d-1} \cdot \left(\frac{1}{2}\right)^d \\
&= c \sum_{d=1}^{d=h} \frac{1}{2} \\
&= c \frac{h\left(1 + \frac{h-1}{2}\right)}{2} \\
&= c \frac{h\left(1+h\right)}{4}
\end{aligned}$$

which of course goes to $\infty$ as $h \to \infty$

### 1.3

Let $p \geq \frac{1}{2}, c > 0$.

#### 1.3.1 Inside Left Subtree

The number of possible edges in the left subtree is

$$\#_L = \binom{\frac{n}{2}}{2} \approx \frac{n^2}{8}$$

Now, for a given node, the numder of edges it can have through a parent at height $k$ is

$$2^{k-1}$$

and we have a total of $\frac{n}{2}$ nodes, so given a tree of height $h$, the total amount of edges using a parent at height $k$ is given by (dividing by 2 to account for duplicate edges):

$$2^{k-1} \cdot \frac{n}{4} = 2^{k-1} 2^{h-2} = 2^{h+k-3}$$

Each realized edge can be seen as a 'successful' result in a bernoulli trial, so the expected number of edges is

$$\sum_{k=1}^{h-1} 2^{h+k-3} \cdot c \cdot p^k = 2^{h-3} \cdot c \sum_{k=1}^{h-1} (2p)^k = 2^{h-3} \cdot c \cdot \frac{2p\left((2p)^{h-1} - 1\right)}{2p-1}$$

which means that the expected density is

$$\mathbb{E}\left(Density\right) = \frac{2^{h-3} \cdot c \cdot \frac{2p\left((2p)^{h-1}-1\right)}{2p-1}}{\frac{n^2}{8}} = 2^{-h} \cdot c \cdot \frac{2p\left((2p)^{h-1}-1\right)}{2p-1}$$

#### 1.3.2 Between the subtrees

Now, the number of potential edges is

$$\#_{Between} = \frac{n^2}{4}$$

Since all of the edges connect the two subtrees, the common ancestor is always the root, therefore the expected number of realized edges is

$$\frac{n^2}{4} \cdot c \cdot p^h$$

and the density is

$$\mathbb{E}\left(Density\right) = \frac{\frac{n^2}{4} \cdot c \cdot p^h}{\frac{n^2}{4}} = c \cdot p^h$$

#### 1.3.3 Clustering

We'll compare the two densities:

$$\frac{D_{Left}}{D_{Between}} = \frac{\left((2p)^{h-1}-1\right)}{(2p-1)\,(2p)^{h-1}} \approx \frac{1}{(2p-1)}$$

and so, as $p \to 1$ the densities grow alike and no clustering occurs, but as $p \to \frac{1}{2}$ then $D_{Left} \gg D_{Between}$ and the two groups form distinct clusters

# 2

Throughout this question we'll follow the same reasoning used for the proofs in the case where $d = 2$, and try and keep notation a close as the origin for brevity and clarity

## 2.1

Let $r = 0 \Rightarrow P\left(u \to v\right) = \frac{1}{n-1}$. Let $N$ denote the set of nodes that reside within a $\sqrt{n}$ radius of the last node (note that $|N| = \sqrt{n}$). The probability of a node $u$ to have a long link into $N$ is:

$$P\left(u \to N\right) = \frac{\sqrt{n}}{n-1} \approx \frac{1}{\sqrt{n}}$$

now, let $\epsilon$ denote the event that in the first $\sqrt{n}$ steps we have a long edge into $N$:

$$\begin{aligned}P(\epsilon) &= 1 - P(\bar{\epsilon}) \\ &= 1 - \left(1 - \frac{\sqrt{n}}{n-1}\right)^{\sqrt{n}} \\ &\approx 1 - e^{-\frac{n}{n-1}} \\ &\approx 0.62\end{aligned}$$

Note that $\epsilon$ is the event that we will take less than $\sqrt{n}$ steps to reach our target. This means that the expected number of steps required:

$$\mathbb{E}(\#_{steps}) > P(\bar{\epsilon}) \cdot \sqrt{n} > 0.37\sqrt{n}$$

## 2.2

Let $r = 1$, then the probability of two node being connected by a long edge is

$$P(u \to v) \frac{d(u,v)^{-1}}{\sum_{w \neq u} d(u,w)^{-1}}$$

denote

$$Z = \sum_{w \neq u} d(u,w)^{-1}$$

and we'll try to bound the value of $Z$.

$$\begin{aligned}Z &\le \sum_{i=1}^{n-1} 2 \cdot i^{-1} \\ &\le 2 \int_{i=1}^{n-1} \frac{1}{i} di \\ &\le 2\left[1 + \ln(n)\right] \\ &\le 2\ln(3n) \\ \Rightarrow Z &\le 2\ln(3n) \qquad (1)\end{aligned}$$

Now, we say that our algorithm is in phase $j$ if we are in node $n$ and $2^j < d(u,n) \le 2^{j+1}$. Initialy, we start at phase $j = \log(n)$, and by design of our algorithm, we progress in a non-decreasing approach w.r.t phases.

Let $X_j$ be an RV that indicates the total number of steps spent at phase $j$, and let $X$ denote the total steps spent until the target is reached; clearly $X = \sum_{j=1}^{\log(n)} X_j$.

Let $L_j$ denote the set of all nodes of distance up to $2^j$ of the target. The distance between the node $v$ in phase j and an arbitrary node in $u \in B_j$ is bounded by:

$$d(v,u) \le d(v,t) \le 2^{j+1}$$

The number of nodes in $L_j$ is bounded by (remember we are talking about a straight line ending at $n$)

$$|L_j| \geq 2^j$$

A long range contact therefore reaches $L_j$ with probability of at least

$$\frac{|L_j| \cdot d\left(v, L_j\right)^{-1}}{Z} \geq \frac{2^j \left[2^{j+1}\right]^{-1}}{2 \ln\left(3n\right)} = \frac{1}{4 \ln\left(3n\right)}$$

Finding a long range contact is therefore a Bernoulli RV with $p \geq \frac{1}{4 \ln(3n)}$, therefore

$$\mathbb{E}\left(X_j\right) = \sum_{i=1}^{\infty} P\left(X_j \geq i\right) \leq \sum_{i=1}^{\infty} P\left(1 - \frac{1}{4 \ln\left(3n\right)}\right)^{i-1} = 4 \ln\left(3n\right)$$

$$\Rightarrow \mathbb{E}\left(X\right) = \sum_{j=1}^{\log(n)} \mathbb{E}\left(X_j\right) \leq 4 \ln\left(3n\right) \log\left(n\right) \tag{2}$$

## 2.3

Let $r = 2$. We'll show that the expected time spent will be at least $a \cdot n^b$. Our probability function now is

$$P\left(u \to v\right) \frac{d\left(u, v\right)^{-2}}{\sum_{w \neq u} d\left(u, w\right)^{-2}}$$

and note that

$$\sum_{w \neq u} d\left(u, w\right)^{-2} \geq 1$$

Let us show that the probability of a link that goes a long way beyond a distance $m$ is low. For a given long distance edge $(u, v)$:

$$\begin{aligned} P\left(d\left(u, v\right) > m\right) &\leq \sum_{j=m+1}^{n} 2 \cdot \frac{j^{-2}}{Z} \\ &\leq \frac{2}{Z} \cdot \sum_{j=m+1}^{n} j^{-2} \\ &\leq 2 \int_{m}^{\infty} j^{-2} dj \\ &= 4 \frac{1}{m} \end{aligned}$$

Now, consider an algorithm that travels through $\frac{1}{8}\sqrt{n}$ new nodes. Denote by $\epsilon_i$ the event in which the algorithm encounters a long edge that leads from the $i_{th}$ node visited a distance of over $\sqrt{n}$. Denote by $\epsilon$ the event that some such long distance edge is encountered in the first $\frac{1}{8}\sqrt{n}$ nodes

$$P(\epsilon) \leq \sum_{i=1}^{\frac{1}{8}\sqrt{n}} P(\epsilon_i) \leq \frac{1}{8}\sqrt{n} \cdot \frac{4}{\sqrt{n}} = \frac{1}{2}$$

Recall that the event $\bar{\epsilon}$ means that we did not encounter long edges in our first $\frac{1}{8}\sqrt{n}$ nodes, so we will travel at least that many nodes to reach our target, meaning that the expected number of steps

$$\mathbb{E}(\#_{steps}) \geq \mathbb{E}(\#_{steps}|\bar{\epsilon})\, P(\bar{\epsilon}) \geq \frac{1}{16}\sqrt{n}$$

# 3

## 3.1

Let $G$ be an undirected connected graph. Let $p$ be the probability distribution vector such that

$$p_i = \frac{\deg(i)}{\sum_{v\in V}\deg(v)} = \frac{\deg(i)}{|E|}$$

We'll now show that $p$ is a stationary distribution.

$$\begin{aligned} p^{new} &= T \cdot p \\ \Rightarrow p_i^{new} &= \sum_{v\in V} T_{i,v} \cdot p_v \\ &= \frac{1}{|E|}\sum_{v\in V} T_{i,v} \cdot \deg(v) \\ &= * \end{aligned}$$

and since we know that

$$T_{i,j} = \begin{cases} \frac{1}{\deg(j)} & (j,i) \in E \\ 0 & (j,i) \notin E \end{cases}$$

then

$$\begin{aligned} * &= \frac{1}{|E|}\sum_{v:(i,v)\in E} \frac{1}{\deg(v)} \cdot \deg(v) \\ &= \frac{\deg(i)}{|E|} \\ \Rightarrow p_i^{new} &= p_i \end{aligned}$$

which means that indeed, $p$ is a stationary distribution.

### 3.2

We will show that in some graphs the second step results in a higher expected value, and in some a lower one.

#### 3.2.1 Higher

Examine the following graph

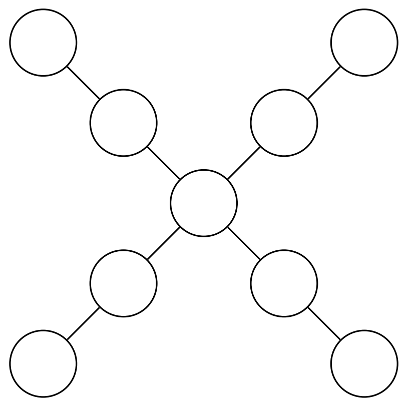

Figure 1: Second Step Raises Expectation

Initially assigning each node $\frac{1}{9}$ we achieve the following results

$$\begin{aligned}\mathbb{E}\left(S_1\right) &= 4\left(\frac{1}{2}\cdot\frac{1}{9}\right)\cdot 1 + 4\left(\frac{1}{9}+\frac{1}{36}\right)\cdot 2 + \left(\frac{1}{2}\cdot\frac{4}{9}\right)\cdot 4\\ &= \frac{20}{9}\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left(S_2\right) &= 4\left(\frac{1}{2}\cdot\frac{1}{9}+\frac{1}{4}\cdot\frac{1}{2}\cdot\frac{1}{9}\right)\cdot 1 + 4\left(\frac{1}{2}\cdot\frac{1}{9}+\frac{1}{2}\cdot\frac{1}{9}\right)\cdot 2 + \left(\frac{1}{2}\cdot\frac{4}{9}+\frac{1}{2}\cdot\frac{1}{9}\right)\cdot 4\\ &= \frac{5}{18}+\frac{16}{18}+\frac{20}{18} = \frac{41}{18}\end{aligned}$$

$$\Rightarrow \mathbb{E}\left(S_2\right) > \mathbb{E}\left(S_1\right)$$

#### 3.2.2 Lower

Examine the following graph

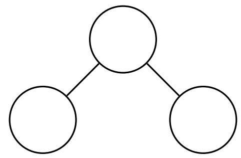

Figure 2: Second Step Raises Expectation

And assign initial values of $\frac{1}{3}$, then

$$\begin{aligned}\mathbb{E}\left(S_1\right) &= 2 \cdot \left(\frac{1}{2} \cdot \frac{1}{3}\right) \cdot 1 + \left(\frac{2}{3}\right) \cdot 2 \\ &= \frac{5}{3}\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left(S_2\right) &= 2 \cdot \left(\frac{1}{2} \cdot \frac{2}{3}\right) \cdot 1 + \left(\frac{1}{3}\right) \cdot 2 \\ &= \frac{4}{3} \\ &\Rightarrow \mathbb{E}\left(S_2\right) < \mathbb{E}\left(S_1\right)\end{aligned}$$

## 3.3

### 3.3.1 Adding Edges

Examine the following graph

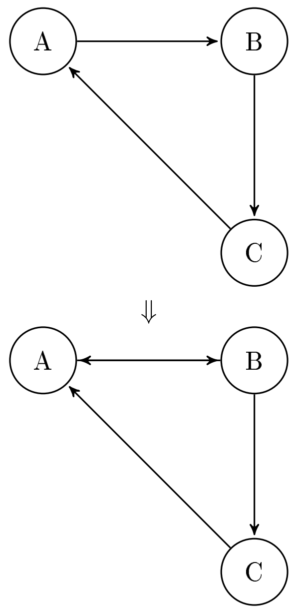

Figure 3: Adding Edge To Increase Value

First note that the first graph will forever stay in the initial distribution ($p^\star = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$). Now, after adding the edge $B \to A$, we get the following transition matrix:

$$T = \begin{bmatrix} 0 & \frac{1}{2} & 1 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{bmatrix}$$

and after some algebra we find that the normalized eigenvector with eigenvalue of 1 is $p^\star = \begin{bmatrix} 0.4 \\ 0.4 \\ 0.2 \end{bmatrix}$, and clearly $B$ has increased the probability of it being visited

#### 3.3.2 Removing Edges

Examine the following diagrams

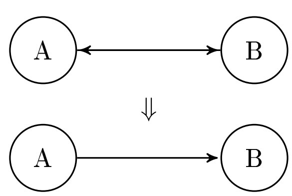

Figure 4: Removing Edges To Increase Value

Again, the original state was balanced and each node had equal chances of being visited. Once removing the $B \to A$ edge, $B$ becomes a sink and will drain all probability towards it thus achieving the desired effect.