Handouts #1 – Proof of Small World Navigation

Instructor: *A. Chaintreau,* Teaching Assistant: *Emily Hua, Ana Stoica*

This handouts contains a full detailed proof of the small world navigation with dimension $k = 1$. It is given as a companion to the first assignment, which generalizes this proof to dimension $k > 1$.

Do not let the relative length discourage you, in most books it is much shorter, but this is precisely why this version is much easier to read!

Definition: A *random biased augmented lattice* of dimension k containing N nodes with bias parameter r is defined as follows:

- We assume $V = \left\{ \ (i_1, \ldots, i_k) \in \{1, 2, \ldots, L\}^k \ \right\}$, (note that $N = L^k$).

- Nodes are connected to all other nodes whose distance in the lattice is at most p (i.e. $v = (i_1, \ldots, i_k)$ and $v' = (i'_1, \ldots, i'_k)$ are connected if $|i_1 - i'_1| + \ldots + |i_k - i'_k| \le p$).

- In addition, each node is connected to q others nodes chosen independently such that

$$\mathbb{P}\left[u \rightsquigarrow v \right] = \frac{\frac{1}{\|u-v\|^r}}{\sum_{w \neq u} \frac{1}{\|u-w\|^r}} \ .$$

The distance p and the number of random shortcuts q are two parameters of the model, which have little effect on the performance of distributed algorithm. We always assume $p = q = 1$. Note that, in the probability describing the chance to connect u and v, the denominator only plays the role of a normalizing constant. For the sake of this handout, we assume $k = 1$.

Theorem 1.
- *When $r = 1$, greedy routing uses in expectation at most $O(\ln(N)^2)$ of steps.*

- *When $0 \le r < 1$, for any p and q, then as n grows any decentralized algorithm uses in expectation at least $\Omega(N^{\frac{1-r}{2}})$*

- *When $r > 1$, for any p and q, then as n grows any decentralized algorithm uses in expectation at least $\Omega(N^{\frac{r-1}{r}})$*

Proof. The proof of each case contains two parts: First, a bound on the normalizing constant used in the probability distribution of the shortcuts. Second, a study of the progress of greedy routing which uses this bound.

Let us first observe the following bounds on the normalizing constant which holds for the case $k = 1$:

$$\sum_{j=1}^{\lfloor N/2 \rfloor - 1} \frac{1}{j^r} \le \sum_{v \neq u} \frac{1}{\|u-v\|^r} \le 2 \sum_{j=1}^{N} \frac{1}{j^r} \ . \tag{1}$$

It can be deduced as follows. First, wherever u is positioned in the line, it has at least one side (either left or right) which contains at least $N/2$ neighbors. For each value of $j = 1, \ldots, \lfloor N/2 \rfloor - 1$, it has

one neighbor at distance j on this side of the line, which proves the lower bound. The upper bound is obtained after observing that u has at most 2 neighbors at distance j for all j and that the maximum distance cannot be more than N.

Intuitively, these inequalities indicate that the behavior of the normalizing constant is closely coupled with the power series with coefficient r. Note that this series is convergent if and only if $r > 1$, which intuitively explain that $r = 1$ is a critical value for the system.

The case $r < 1$ For $r < 1$, a similar argument as the one used to study uniformly augmented lattice holds. As we wish to show a similar negative result, our goal is to show that the probability to find a shortcut is "small" and hence to find a lower-bound for the normalizing constant. By Eq.(1), we have

$$\sum_{v \neq u} \frac{1}{\|u - v\|^r} \geq \sum_{j=1}^{\lfloor N/2 \rfloor - 1} \frac{1}{j^r} \geq \int_1^{\lfloor N/2 \rfloor} \frac{1}{x^r} dx \geq \frac{1}{1-r} \left((\lfloor N/2 \rfloor)^{1-r} - 1 \right)$$

The second inequality comes from the fact that, as $x \mapsto \frac{1}{x^r}$ is a decreasing function, it is smaller than $\frac{1}{j^r}$ on the interval $[j, j+1[$. The last inequality simply follows from computing the integral.

As a consequence, the sum used in the normalizing constant asymptotically grows polynomially, with coefficient $1 - r > 0$. In particular, for $N \geq 2^{\frac{2-r}{1-r}}$ we have that $(N/2)^{1-r} \geq 2$ and hence $(N/2)^{1-r} - 1 \geq \frac{1}{2}(N/2)^{1-r}$. We then deduce:

$$\text{For } N \geq 2^{\frac{2-r}{1-r}}, \sum_{v \neq u} \frac{1}{\|u-v\|^r} \geq c_1 N^{1-r} \text{ where } c_1 = \frac{1}{2(1-r)2^{(1-r)}} .$$

This proves that, however u and v are located, the probability that the shortcut originating in u leads to v is becoming small *polynomially* with N:

$$\mathbb{P}[u \rightsquigarrow v] \leq \frac{1}{c_1 N^{1-r}}$$

This is sufficient for the proof used on uniformly augmented lattice to apply: If we denote again by I_l the set of nodes at distance at most l from the target,

$$I_l = \{ u \in V \mid |u - t| \leq l \} ,$$

then since the number of nodes in this subset is less than $2l$, the probability that a shortcut originated in u leads to a node in I_l is upper bounded by $\frac{2l}{c_1 N^{1-r}}$.

We may now consider the sequence of nodes visited by the greedy routing procedure $U_1, U_2, \ldots, U_k$, and for each of them denote by X_i the destination of the shortcut originating at U_i. The probability that one of the n first elements of X_i lies in I_l is then upper bounded by the union bound:

$$\mathbb{P}\left[\bigcup_{i=1,\ldots,n} \{ X_i \in I_l \} \right] \leq \sum_{i=1,\ldots,n} \mathbb{P}[X_i \in I_l] \leq \frac{n2l}{c_1 N^{1-r}} .$$

Note the similarity with the proof for the random uniform augmentation, with the only difference being a constant and a different coefficient for the power of N in the denominator.

Choosing $l = n = \lambda N^{\frac{(1-r)}{2}}$ the probability above is upper bounded by a constant independent of N. By choosing λ sufficiently small we have that it is less than $1/4$.

This indicates that with probability at least $3/4$ all the n first shortcuts found by greedy routing connect with a node outside of I_l. On this probability event, we know that starting from a point s outside of I_l, greedy routing cannot succeed in finding the target t in less than $\min(n, l) = \lambda N^{\frac{1-r}{2}}$. Indeed, finding the target requires here either to use more than n steps or to traverse from the border of I_l to the target using only local edges.

In this theorem, we analyze the *expected* performance of greedy routing starting from an initial random point. With a probability at least $1/2$ the distance between this node s and t is no less than $n/4$, which implies for sufficiently large n that s does not start in I_l. We can then state that, in expectation, greedy routing needs a number of steps at least $(1/2)(3/4)\lambda N^{\frac{1-r}{2}}$.

The case $r > 1$: As in the previous case, we wish to establish a negative result hence our goal will be to provide an upper bound on the chance to make sufficient progress. However, the argument will be different this time, as the main obstacle is that the probability of having sufficiently long shortcuts is not large enough to allow the greedy procedure to move towards the destination sufficiently fast.

Indeed, we know that any node u in the line has at most 2 neighbors with distance j in the lattice, and the series which characterizes the normalizing constant, as shown in Eq.(1) converges. The fact that the series converge indicates that we deduce now a bound on the probability of reaching all nodes that are sufficiently far.

$$\sum_{v \neq u, \|u-v\| > m} \frac{1}{\|u-v\|^r} \leq 2 \sum_{j=m+1}^{N} \frac{1}{j^r} \leq 2 \left(\int_m^N \frac{1}{x^r} dx \right) \leq \frac{2}{(r-1)m^{r-1}} \ .$$

The last inequality is obtained after replacing the integral on $[m; N[$ with the integral on $[m, +\infty[$, which can only make this bound looser, and computing its value.

Since the normalizing constant is always greater than 1 the inequality above implies that for any m the probability for any node u to be connected through a shortcut to a node at distance larger than m is less than $\frac{2}{(r-1)m^{r-1}}$. We now consider the n first shortcuts encountered by greedy routing, as made in the previous proof. Following the union bound, we can deduce that the probability that at least one of them connect two nodes at distance larger than m is less than n times the above probability (*i.e.* $\frac{2n}{(r-1)m^{1-r}}$).

Let us now assume that we can choose m and n in such a way that this probability is smaller than $1/4$, this would implies that with probability at least $3/4$ all the n first encountered shortcuts connect two nodes at distance at most m (a probability event we denote by $\mathcal{E}$). We may assume that the initial distance between s and t is at least $N/4$. This event occurs with a probability $1/2$ and hence in intersects the event $\mathcal{E}$ at least for a probability $1/4$.

When both event occur then in order to complete the walk from s to t, greedy routing requires at least $\min(n, N/4m)$ steps, since the first n steps of the walk has a length at most m. This would imply, in expectation, that number of steps needed by greedy routing is at least $\frac{1}{4}\min(n, N/4m)$.

Now in order to complete the proof, we need to show that we can choose n and m so that $\frac{2n}{(r-1)m^{r-1}} \leq 1/4$ and $1/4 \min(n, N/4m)$ is large as N grows.

The first condition is satisfied as long as $n \leq \frac{2}{4(r-1)}m^{r-1}$. We can choose n to be exactly this value as making n large is only helping to satisfy the second condition. Hence, both conditions reduces to finding m such that $\min(\frac{2}{4(r-1)}m^{r-1}, N/4m)$ is large as N grows.

The value of m has opposite role in order to maximize each term, so that intuitively this minimum will be the largest when the two terms have the same order. In particular, if we choose $m = N^{\frac{1}{r}}$ this minimum is a constant multiplied by $N^{\frac{r-1}{r}}$, which proves the result of the theorem.

The case r=1: Finally we are left with the only positive result, which occurs at the critical case. It is interesting to see first why the proof of the case $r < 1$ and $r > 1$ do not apply. First, for $r = 1$ the series characterizing the normalizing constant is the harmonic series, hence it does not converge and we cannot apply the previous argument bounding the probability to find large links. Also, as opposed to the case $r < 1$ the series does not grow as fast as a polynom, which explains why we cannot use this argument to show that all probability, independently of the position of u and v becomes small.

We first obtain a upper bound on the series, as we observed that it diverges not as fast as polynom. Indeed, when $r = 1$, following Eq.(1),

$$\sum_{v \neq u} \frac{1}{\|u - v\|} \leq 2(1 + \sum_{j=2}^{N} \frac{1}{\|j\|}) \leq 2(1 + \int_{1}^{N} \frac{1}{x} dx) \leq 2(1 + \ln(N)) \leq 2(\ln(3N)) \ .$$

This implies that for any u the probability that it it connected with a shortcut to node v is at least $1/(2 \ln(3N) d(u,v))$.

Greedy routing, initially started in a point s constructs a chains of nodes visited $U_1, U_2, \ldots$ until it reaches t. Let us say that U_i is in phase j if we have $2^j \leq \|U_i - t\| \leq 2^{j+1}$. Since the initial distance is at most N, we now that U_1, the starting point of the walks is in phase j_0 with $j_0 \leq \ln(N)/\ln(2)$. Note also that, as greedy routing decreases the distance to the target at each step, the phase of this walk can only decreases with the number of steps made.

The core of the argument for the theorem is to show that each phase of this walk is short (*i.e.* it involves a logarithmic number of step). This will imply the result because there are also a small number of phase (*i.e.* a logarithmic number).

We first consider the following quantity: Given that U_i is in phase j, what is the probability that U_{i+1} is in phase $j' < j$? According to the definition, all nodes in phase $j' < j$ are those who are at distance at most 2^j from the target t. This contains at least 2^j nodes (the target t may be on the border of the line, but it has at least 2^j neighbors within this distance in one direction).

The key observation is that, for every node v in phase $j' < j$, since U_i is in phase j, the distance between U_i and v can be bounded by triangular inequality:

$$\|U_i, v\| \leq \|U_i, t\| + \|t, v\| \leq 2^{j+1} + 2^j \leq (3/2)2^{j+1} \ .$$

Hence, the probability that U_i has a shortcut leading to a node in phase $j' < j$ is at least

$$\sum_{v \mid \|v - t\| < 2^j} \frac{1}{2 \ln(3N)(3/2)2^{j+1}} \geq \frac{2^j}{(2 \ln(3N)(3/2)2^{j+1})} \geq \frac{1}{6 \ln(3N)} \ .$$

In other words, for any step taken by greedy routing in phase j the next step will be in a smaller phase with probability at least $(6 \ln(3))^{-1}$. Note that this event only depends on the shortcuts chosen at this step and hence, the shortcuts that will be visited in the next steps are independent from this event.

This implies that, if we denote by S_j the number of steps made by this walk inside phase j, we can bound the probability that $S_j \geq i$ geometrically, hence we have:

$$\mathbb{E}[S_j] = \sum_{i \geq 1} \mathbb{P}[S_j \geq i] \leq \sum_{i \geq 1} \left(1 - \frac{1}{6 \ln(3N)}\right)^{i-1} = 6 \ln(3N).$$

Assuming that greedy routing starts in phase j_0, the total number of steps it needs to reach t is $S_{j_0} + S_{j_0-1} + \ldots + S_1$. By the linearity of expectation, and since $j_0 \leq \ln(N)/\ln(2)$, we have that it takes in expectation less than $6/(\ln(2)) \ln(3N) \ln(N)$, which is less than $c \ln^2(N)$ for some constant c, proving the result.