FS 2013 Prof. R. Wattenhofer / Michael König

Principles of Distributed Computing
Exercise 8: Sample Solution

1 Diameter of the Augmented Grid

a) As our target node set is of size $\Omega(\log^2 n)$ and link targets are distributed uniformly at random over all n nodes, each link connects to the target set with probability $p \in \Omega((\log^2 n)/n)$. Thus, for sufficiently large[1] n, the probability that $n/\log n$ many links miss the set is bounded by

$$(1-p)^{n/\log n} \leq e^{-pn/\log n} \in e^{-\Omega(\log n)} = \frac{1}{n^{\Omega(1)}}.$$

Now we exploit the power of the Big-O notation. Choosing a sufficiently large multiplicative constant in front of the $(n/\log n)$-term, this becomes a bound of $1/n^c$, and choosing a large additive constant, we make sure that the bound holds also for the values of n that are not "sufficiently large". Thus, the probability that at least one link enters the set of $\Omega(\log^2 n)$ nodes is at least $1-1/n^c$, i.e., this event occurs w.h.p..

In order to obtain the same result using a Chernoff bound, let X_i, $i \in \{1,\ldots,l\}$, where $l = c_1 \cdot n/\log n$ (for an arbitrarily large constant c_1) is the number of considered links, be random variables that are 1 if the i^{th} link ends in the set (i.e., with the probability p from above) and 0 otherwise. Defining $X := \sum_{i=1}^{l} X_i$, we get that $\mathbb{E}[X] = pl$. Plugging in the values yields that $\mathbb{E}[X] \geq c_2 \log n$ (note that c_2 depends on c_1 and hence can also be an arbitrarily large constant).

The Chernoff bound now yields for $\delta = \frac{1}{2}$

$$Pr\left[X \leq \frac{1}{2}\mathbb{E}[X]\right] \leq e^{-\mathbb{E}[X]/8} \leq \frac{1}{n^{c_3}}$$

(for a constant c_3 using that c_2 is an arbitrarily large constant) which means that at least one link points to our target set with high probability.[2]

b) Because $|S| \in o(n)$, also $O(|S|) \subset o(n)$, i.e., the union of the set S ($|S|$ nodes) with the destinations of the $|S|$ random links and all grid neighbors of such nodes (at most $5|S|$ many nodes) has $o(n)$ nodes (because $O(|S|) \subset o(n)$). Thus, always $n - o(n) = (1-o(1))n$ nodes can be found which neither have been visited themselves nor have any neighbors that have been visited so far. Hence, regardless of the choice of the set S and any random links leaving S we have (sequentially) examined up to now, any uniformly independent random choice will contribute 5 new nodes with some probability $p \in \Omega(1-o(1))$.

Let X_i, $i \in \{1,\ldots,|S|\}$, be random variables that are 1 if the random link of the i^{th} node points to a node which neither has been visited itself nor has any grid neighbors that have been visited so far (this happens with probability p), and that are 0 otherwise. Let

[1] This phrase means for some constant n_0, the statement will hold for all $n \geq n_0$.

[2] Small values of n are again dealt with by the additive constant in the O-notation. In general, it is always feasible to assume that n is "sufficiently large" when proving asymptotic statements.

$X := \sum_{i=1}^{|S|} X_i$. The linearity of the expectation value gives us $\mathbb{E}[X] \in \Omega((1-o(1))|S|)$. Now we use a Chernoff bound on the number of "good" choices X and set δ to $\frac{1}{\sqrt{\log n}}$:

$$\Pr[X \leq (1-\delta)\mathbb{E}[X]] \leq e^{-\mathbb{E}[X]\delta^2/2} \leq e^{-\Omega(logn)} \leq \frac{1}{n^c}$$

which yields that the number of "good" choices will be in $(1-o(1))(1-o(1))|S| = (1-o(1))|S|$ w.h.p. (instead of just in expectation). Thus, in total we reach $(5-o(1))|S|$ many nodes w.h.p..

c) Recall that we may choose the constant c in "w.h.p." by ourselves. Thus, we may decide that in a Chernoff bound, it is $c' := c+1$. Hence, the probability that in a given step our set grows by a factor $(5-o(1))$ (provided that $|S| \in o(n)$, as we use part **b)**) is always at least $1 - 1/n^{c'}$. This means in at most a fraction of $1/n^{c'}$ of the events, something goes wrong in a single step. We need less than $\log n$ steps to get to $O(n/\log n)$ nodes, as the number of nodes more than quadruples in each step. In total, in a fraction of less than $\log n/n^{c'} = \log n/n \cdot 1/n^c < 1/n^c$ of all cases something goes wrong. This idea is called union bound.

d) Using the union bound again, we plug together the facts that (i) each node can reach $\Omega(\log^2 n)$ nodes following grid links only within $\log n$ steps, (ii) starting from these nodes, with high probability $O(n/\log n) \subset o(n)$ nodes can be reached within $O(\log n)$ more hops (part **c)**), (iii) from these nodes we reach with high probability the $(\log n)$-neighborhood (with respect to the grid) of any node (part **a)**), and (iv) from there on we can reach the respective node with $\log n$ hops on the grid. Combining this yields that with high probability in total $O(\log n)$ hops are necessary to reach some node v starting at some other node u. Finally, observe that we have $n(n-1) < n^2$ possible (ordered) combinations of nodes; choosing $c' := c+2$ and applying the union bound once more, we infer that we have with high probability a path of length $O(\log n)$ between any pair of nodes, i.e., the diameter of the graph is in $O(\log n)$ w.h.p..

2 Greedy Routing in the Augmented Grid

The key observation is that the probability bound on advancing to the next phase is independent of the current node, as we always inspect fresh random links (recall that we get closer to the destination with each step). In each step, we proceed with probability $p \in \Omega(1/\log n)$ to the next phase. We know that after $\log n$ phases we reach the target. Let $X := \sum_{i=1}^{(C\log n)^2} X_i$, where the X_i are independent $0-1$ variables taking the value 1 with probability p. This can be interpreted as the number of phases we passed in $(C\log n)^2$ steps. Apparently, we can set C such that $\mathbb{E}[X] \geq 2\log n$.

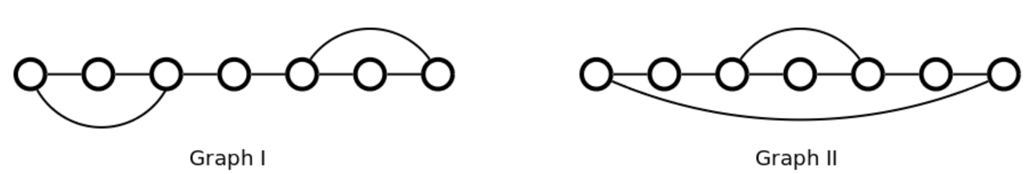
The probability that we use more than $(C\log n)^2$ steps is Pr[number of steps $> (C\log n)^2] \leq \Pr[X \leq \log n]$. Now we can use the Chernoff Bound with $\delta = \frac{1}{2}$ to show that:

$$\Pr[X \leq (1-\delta)\mathbb{E}[X]] = \Pr[X \leq (1-\frac{1}{2})2\log n] = \Pr[X \leq \log n] \leq e^{-\mathbb{E}[X]\delta^2/2} \leq e^{-\Omega(\log n)} \leq \frac{1}{n^c}$$

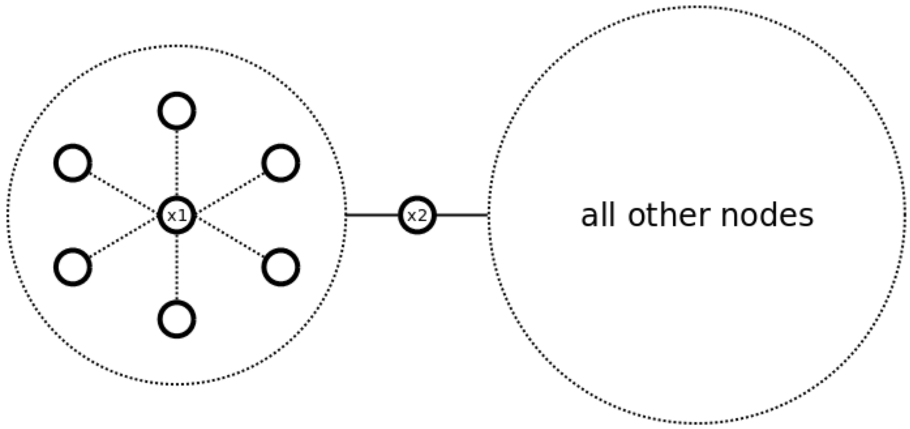
So the probability that we use more than $(C\log n)^2$ steps is smaller than $\frac{1}{n^c}$, which means greedy routing terminates after $O(\log^2 n)$ steps w.h.p..

3 Scale Free Networks

a) The node-degree distribution of a graph does not determine the diameter of the graph. This can be seen in the following figure, where Graph I and Graph II have the same node-degree distribution, but have different diameters: Diameter(Graph I)=4 and Diameter(Graph II)=3.



b) Choosing the node with the highest degree might be a good tactic in general but, there are power law graphs, where the node with the highest degree is not the optimal choice. A sketch of such a network can be seen in the following figure.



In a graph with n nodes and an approximate power law node-degree distribution, the highest degree of a node can be less than $\frac{n}{2}$. In the depicted network node $x1$ is the node with the highest degree and all nodes adjacent to $x1$ are arranged in a star. This star is connected through only one node to the rest of the graph. This means that node $x2$ is closer to more than $\frac{n}{2}$ of the nodes.