Albert-Ludwigs-Universität
Institut für Informatik
Prof. Dr. F. Kuhn July 25, 2012

Network Algorithms, Summer Term 2012
Problem Set 8 – Sample Solution

Exercise 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

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

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 $X_i, i \in \{1, \ldots, \ell\}$, where $\ell \in O(n/\log n)$ 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 $X := \sum_{i=1}^{\ell} X_i$, we get that $\mathbb{E}[X] = p\ell$. Picking a constant $C > 0$ and properly adapting the constants in the $O(n/\log n)$-term, we get that $\mathbb{E}[X] \geq C \log n$. Thus, the Chernoff bounds states that for sufficiently large values of C and n (we need to cope with the fact that we do not know the constants in the O-term in the Chernoff bound), we have that $|X - \mathbb{E}[X]| < C \log n$, w.h.p. Because $\Pr[X > 0] \geq \Pr[|X - \mathbb{E}[X]| < E[X]]$, we know that $X > 0$ also holds w.h.p., 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 $p \in 1 - o(1)$. The linearity of the expectation value gives us $\mathbb{E}[X] \in (1 - o(1))|S|$. Now we use a Chernoff bound (Bound 2) on the number of such "good" choices and set δ to $1/\sqrt{\log n}$:

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

which yields that the number of "good" choices will be in $(1 - o(1))|S|$, w.h.p. (instead of just in expectation).[3] Thus, in total we reach $(5 - o(1))|S|$ many nodes, w.h.p.

[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.

[3] Since the expected value $\mathbb{E}[X]$ of the respective random variable X is large compared to $\log n$ (here we use $|S| \in \Omega(\log^2 n)$), the deviation from the expected value is with high probability in $o(\mathbb{E}[X])$.

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 of $(5 - o(1))$ (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 $(1/n^{c'})$-fraction of the events, something goes wrong in a single step. We need less than $\log n$ steps to get to $O(n/logn)$ 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 argument is also known as a union bound as for any collection of events $A_1, \ldots, A_k$, $\Pr(A_1 \cup \cdots \cup A_k) \leq \Pr(A_1) + \cdots + \Pr(A_k)$.

d) Using a union bound again, we put 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 b)), (iii), from these nodes we reach with high probability the $(\log n)$-grid-neighborhood of any node (part a)), and (iv), from there on we can reach the respective node within $\log n$ 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 a union bound once more, we infer that we have w.h.p. a path of length $O(\log n)$ between any pair of nodes. Hence, the diameter of the graph is $O(\log n)$, w.h.p.