Principles of Distributed Computing
Exercise 11

1 Communication Complexity of Set Disjointness

In the lecture we studied the communication complexity of the equality function. Now we consider the disjointness function: Alice and Bob are given subsets \( X, Y \subseteq \{1, \ldots, k\} \) and need to determine whether they are disjoint. Each subset can be represented by a string. E.g. we define the \( i^{th} \) bit of \( x \in \{0, 1\}^k \) as \( x_i := 1 \) if \( i \in X \) and \( x_i := 0 \) if \( i \notin X \). Now define disjointness of \( X \) and \( Y \) as:

\[
\text{DISJ}(x, y) := \begin{cases} 
0 & \text{there is an index } i \text{ such that } x_i = y_i = 1 \\
1 & \text{else}
\end{cases}
\]

a) Write down \( M^{\text{DISJ}} \) for the \( \text{DISJ} \)-function when \( k = 3 \).

b) Use the matrix obtained in \( a \) to provide a fooling set of size 4 for \( \text{DISJ} \) in case \( k = 3 \).

c) In general, prove that \( CC(\text{DISJ}) = \Omega(k) \).

2 Distinguishing Diameter 2 from 4

In the lecture we stated that when the bandwidth of an edge is limited to \( O(\log n) \), the diameter of a graph can be computed in \( O(n) \). In this problem, we show that we can do faster in case we know that all networks/graphs on which we execute an algorithm have either diameter 2 or diameter 4. We start by partitioning the nodes into sets: Let \( s := s(n) \) be a threshold and define the set of high degree nodes \( H := \{v \in V \mid d(v) \geq s\} \) and the set of low degree nodes \( L := \{v \in V \mid d(v) < s\} \).

Next, we define: An \( H \)-dominating set \( \text{DOM} \) is a subset \( \text{DOM} \subseteq V \) of the nodes such that each node in \( H \) is either in the set \( \text{DOM} \) or adjacent to a node in the set \( \text{DOM} \). Assume in the following, that we can compute an \( H \)-dominating set \( \text{DOM} \) of size \( \frac{n \log n}{s} \) in time \( O(D) \).

a) What is the distributed runtime of Algorithm 2-vs-4 (stated next page)? In case you believe that the distributed implementation of a step is not known from the lecture, find a distributed implementation for this step! **Hint: The runtime depends on \( s \) and \( n \).**

b) Find a function \( s := s(n) \) such that the runtime is minimized (in terms of \( n \)).

c) Prove that if the diameter is 2, then Algorithm 2-vs-4 always returns 2.

Now assume that the diameter of the network is 4 and that we know vertices \( u \) and \( v \) with distance 4 to each other.
Algorithm 1 “2-vs-4”. Input: $G$ with diameter 2 or 4  
Output: diameter of $G$

1: if $L \neq \emptyset$ then
2: choose $v \in L$ \Comment{We know: This takes $O(D)$.}
3: compute a BFS tree from each vertex in $N_1(v)$
4: else
5: compute an $H$-dominating set $DOM$ \Comment{Use: Assumption or Problem 3)}
6: compute a BFS tree from each vertex in $DOM$
7: end if
8: if all BFS trees have depth 2 or 1 then
9: return 2
10: else
11: return 4
12: end if

---

d) Prove that if the algorithm performs a BFS from at least one node $w \in N_1(u)$ it decides “the diameter is 4”.

e) In case $L \neq \emptyset$: Prove that the algorithm either performs a BFS of depth at least 3 from some node $w$. **Hint: use d)**

f) In case $L = \emptyset$: Prove that the algorithm performs a BFS from at least one node in $N(u)$.

g) Give a high level idea, why you think that this does not violate the lower bound of $\Omega(n / \log n)$ presented in the lecture!

h*) Prove or disprove: If the diameter is 2, then Algorithm 2-vs-4 will always compute some BFS tree of depth exactly 2.

### 3 Computation of an $H$-Dominating Set $DOM$

*Solving this problem is optional/voluntary* but helps understanding Chernoff Bounds by using a simplified version (Bound 2 stated in Problem Set 9 when $\delta := 1/2$.) We show that an $H$-Dominating Set $DOM$ (as used in Algorithm 2-vs-4) can be computed fast.

**Theorem 1 (Awesome Chernoff Bounds – again :-)** Let $X := \sum_{i=1}^{N} X_i$ be the sum of $N$ independent $0 - 1$ random variables $X_i$, then $Pr\left[X \leq \frac{1}{2}E[X]\right] \leq e^{-E[X]/8}$.

a) Warm up: Consider $N$ tosses of a perfect coin. Let the random variable $X_i$ be 1 if the $i^{th}$ coin toss results in “head” and let $X_i$ be 0 otherwise. Define $X := \sum_{i=1}^{N} X_i$, compute $E[X]$ and show that $Pr\left[X \leq \frac{N}{4}\right] \leq e^{-N/16}$.

b) Now we get back to our original problem: Assume all nodes know $n$ and $s$. Let each node in $V$ mark itself with probability $\frac{8(c+1) \cdot \ln n}{s}$, where $\ln$ is the *natural logarithm* with base $e$ and $c$ is an arbitrary constant. Let $X_u$ be the random variable indicating whether node $u$ marked itself. That is $X_u := 1$ if $u$ marked itself and $X_u := 0$ in the other case. Define $X^v := \sum_{u \in N(v)} X_u$. Show that if $v \in H$, then $E[X^v]$ is at least $8(c+1) \cdot \ln n$.

c) Using the Chernoff Bound, show that w.h.p. $v \in H$ has at least one marked neighbor. **Hint:** Use $Pr\left[X \leq 4c \cdot \ln n\right] \leq e^{-E[X]/8}$ as an intermediate step.

d) What is the probability that the set $S$ of all marked nodes is a dominating set of $H$? **Hint:** Use $(1 + x/n)^n \geq e^x$ and $e^x \geq 1 + x$.

e) What is the expected size of $S$? Use Chernoff and prove $Pr[|S| \geq 4(c + 1) \cdot \frac{n \ln n}{s}] \geq 1 - 2^{-\Omega(\sqrt{n \ln n})}$.

f) What is the time complexity of computing an $H$-dominating set $DOM$ of size $O(\frac{n \log n}{s})$ when all nodes know $s$ and $n$ and start at the same time?