Exercise 5: Size matters!

Task 1: As small as possible, please? $(3+2+3+3+2+2)$

In this exercise, we will see that MIS and MDS are related problems. Also, we will see how to properly re-use an algorithm that only works w.h.p.

A *forest decomposition* of a graph $G = (V, E)$ is a decomposition of G into directed forests $F_1 = (V, E_1), \ldots, F_f = (V, E_f)$, such that (i) each $e \in E$ occurs in one and only one E_i, and (ii) every $v \in V$ knows, for every forest F_i, its parent node w.r.t. F_i if applicable.

Let M be an MDS of G. Consider the following minimum dominating set (MDS) approximation algorithm. Suppose a given graph $G = (V, E)$ is decomposed into f forests, such that each $v \in V$ initially knows the set $P(v)$ of its (at most) f parents.

Algorithm 1 MDS approximation algorithm based on a forest decomposition.

$H := \left(V, \left\{\{v, w\} \in \binom{V}{2} \mid P(v) \cap P(w) \neq \emptyset\right\}\right)$
compute an MIS I of H
$D' := \bigcup_{v \in I} P(v)$
$D := D' \cup \{v \in V \setminus D' \mid v \text{ has no neighbor in } D'\}$
return D

a) Show that Algorithm 1 can be implemented in the synchronous message passing model (i.e., LOCAL) with running time $\mathcal{O}(\log n)$ w.h.p.![1]

b) Denote by $V_C \subseteq V$ the set of nodes that are in M or have a child in M. Show that $|V_C| \leq (f+1)|M|$!

c) Denote by $V_P \subseteq V$ the set of nodes that have some parent in M. Show that $|I \cap V_P| \leq |M|$!

d) Prove that at most $(f+1)|M|$ nodes are not covered by D'.

e) Conclude that Algorithm 1 computes a dominating set that is at most by factor $\mathcal{O}(f^2)$ larger than the optimum!

   Hint: $V = V_C \cup V_P$.

f)* Show that even if we restrict message size to $\mathcal{O}(\log n)$ bits, the algorithm can be implemented with running time $\mathcal{O}(\log n)$ w.h.p.

Task 2: Lots of Wood $(3+3+3+3+2)$

Denote by $A(G)$ the *arboricity* of $G = (V, E)$, i.e., the minimum number of forests into which E can be decomposed. Our goal in this exercise is to decompose G into $f \in \mathcal{O}(A)$ forests.

a) Show that in each iteration of the WHILE loop, at least half of the remaining nodes are deleted!

   Hint: Assume that this is false and bound the number of remaining edges from below. Compare the result to the maximum number of edges in $A(G)$ forests.

[1] You may want to look at the appendix, specifically Definition A.5, "With High Probability"

Algorithm 2 Forest decomposition, initial $A(G)$ is known.

$A_{init} := A(G)$
while $V \neq \emptyset$ **do**
  for all $v \in V$ with $\delta_v \leq 4A_{init}$ in parallel **do**
    assign neighbors as parents in forests $F_1, \ldots, F_{4A_{init}}$; break ties by node id
    delete v (and its incident edges) from G
  end for
end while
return the computed forests (each node knows its parent in $F_1, \ldots, F_{4A_{init}}$)

b) Conclude that the algorithm computes a decomposition of G into at most $4A(G)$ directed forests in $\lceil \log n \rceil + 1$ rounds!

c) Change the algorithm so that it does not require knowledge of $A(G)$, but instead relies only on n! You may use up to $8A(G)$ forests and increase the running time of the algorithm by a factor of $\mathcal{O}(\log A(G))^2$.

Hint: Forest decomposition is greedy.

d) Conclude that in graphs of arboricity A, a factor-$\mathcal{O}(A^2)$ approximation to MDS[3] can be found in $\mathcal{O}(\log n \log A)$ rounds w.h.p., even if n and $A(G)$ are unknown, but an upper bound $N \geq n$ is given, with $N \in n^{\mathcal{O}(1)}$!

e)* Can you do it in only $\mathcal{O}(\log n)$ rounds, only with $N \geq n$, $N \in n^{\mathcal{O}(1)}$ known?

Hint: Exploit the upper bound $A(G) \leq n - 1$ and the fact that we use the LOCAL model.

Task 3*: Exponential Enhancement $(1 + 1 + 2 + 2 + 1 + 1)$

a) Why is Chernoff's bound called Chernoff's bound?

b) Show that for independent variables X_i, $i \in I$, $\mathbb{E}\left[\prod_{i \in I} X_i\right] = \prod[\mathbb{E}[X_i]]$.

c) Let X_i, $i \in I$, be random variables, and define $X = \sum_{i \in I} X_i$. Use Markov's bound to show that for arbitrary $t, \delta > 0$,

$$P[X \geq (1+\delta)\mathbb{E}[X]] \leq \frac{\mathbb{E}\left[\prod_{i \in I} e^{tX_i}\right]}{e^{t(1+\delta)\mathbb{E}[X]}}.$$

d) Use b) and c) to infer that if the X_i are independent Bernoulli variables, then

$$P[X \geq (1+\delta)\mathbb{E}[X]] \leq \frac{e^{(e^t-1)\mathbb{E}[X]}}{e^{t(1+\delta)\mathbb{E}[X]}}.$$

e) Plug in $t := \ln(1+\delta)$. You obtain the upper tail bound; choosing $\delta \in (0,1)$ and $t = 1 - \delta$ yields the lower tail bound.[4] The bounds derived here are stronger than those in the lecture, but more unwieldy. For most applications, the simpler versions suffice.

f) Enlarge the knowledge of the exercise group by reporting your findings!

[2] Forest decompositions into f forests are particularly interesting if $f \geq A(G)$ is small, hence usually $\log A(G)$ is *very* small!

[3] Read: "a dominating set at most factor $\mathcal{O}(A^2)$ larger than an MDS."

[4] Note that one has to introduce a minus sign in the exponents in b) to still be able to apply Markov's inequality.