6.S899 Distributed Graph Algorithms Lecture 5 MIT 10/10/2014

Material covered:

Definition of the CONGEST model, the GHS-Algorithm to compute MST and a faster version by Garray, Kutten and Peleg (with a simplified analysis at the cost of the time bound). These can be found (with better time bounds) in Sections 5.5 and 24.2 which can be found in the book “Distributed Computing” by David Peleg, SIAM Monographs on discrete mathematics and applications, volume 5, 2000.

We extended this version of the Garray-Kutten-Peleg algorithm to achieve $\tilde{O}(\sqrt{n}+D)$ rounds. The basic idea of keeping the diameter of fragments below $\tilde{O}(\sqrt{n})$ follows “Shay Kutten and David Peleg, Fast distributed construction of k-dominating sets and applications, Proceedings of the fourteenth annual ACM symposium on Principles of distributed computing (PODC’95), pages 238-251, 1995”. However, we do not use their construction of k-dominating sets, but apply the Cole-Vishkin Algorithm to compute MIS covered in lecture 1. This is inspired by the use of matchings such as done e.g. in “Christoph Lenzen and Boaz Patt-Shamir, ACM Symposium on Principles of Distributed Computing (PODC'14), pages 262-271, 2014” and we did a few simplifications at the cost of log-factors in the runtime.

LAST LECTURE: (C,D)-NETWORK DECOMPOSITION

DECOMPOSE V INTO DISJOINT CLUSTERS $A_1, A_2, \ldots$ S.t.

1) $\text{diam}(A_i) \le D$

2) CLUSTER-GRAPH IS C-COLORABLE

THIS LECTURE: MINIMUM SPANNING TREE

LOCAL MODEL $\square$ $O(D)$

CONGEST MODEL $\square$ $O(nD)$ $\square$ $O(|E|)$

$\square$ $O(n \log n)$

$\square$ $O(n^{0.775} \log n + D \log n)$

$\square$ $O(\sqrt{n} \log^2 n \log^* n + D \log n)$

Def: MST. GIVEN $G=(V,E,\omega)$, $G' \subseteq G$, THEN

$$\omega(G') := \sum_{e \in E'} \omega(e)$$

ST T IS AN MST OF G IF

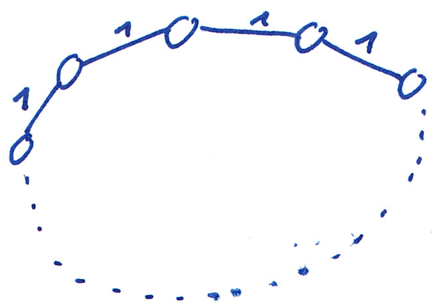
$$\omega(T) = \min \{ \omega(T') \mid T' \text{ IS ST OF } G \}$$

ALGO (LOCAL MODEL): NODE 1
- COLLECTS G'S TOPOLOGY
- COMPUTES MST T
- BROADCASTS T TO G

THM: TAKES TIME $O(D)$ AND THIS IS OPTIMAL.

PROOF: $O(D)$: ✓

$\Omega(D)$:



WHICH EDGE NOT TO PUT INTO T?
$\rightsquigarrow$ TAKES $\Omega(D)$ ROUNDS.

$\square$

DEF: **CONGEST(B) MODEL**

LIKE LOCAL MODEL, BUT MESSAGES HAVE SIZE B (BANDWIDTH B CAUSES CONGESTION)

TYPICAL $B = O(\log n) \approx$ ENCODES CONSTANT MANY IDS

↳ CONGEST MODEL

NOTE: ALGO I TAKES TIME $O(|E|)$ IN THE CONGEST MODEL:

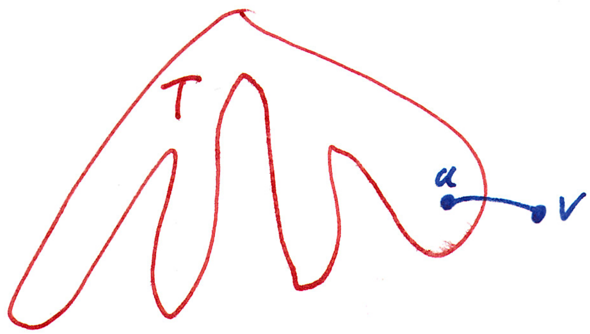
IN THE WORST CASE

NODE 1 NEEDS TO COLLECT $|E|$ EDGES THROUGH THIS EDGE.

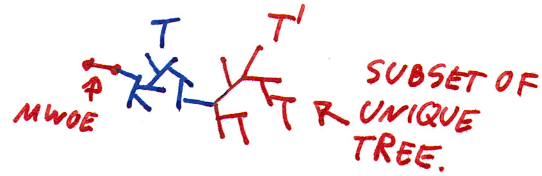
ASSUME: ALL WEIGHTS DISTINCT. $\Rightarrow$ MST UNIQUE
(NO CYCLES TO BREAK, USE IDS ...)

DEF: T IS A MST-FRAGMENT OF G IF THERE EXISTS A MST T_{MST} OF G s.t. $T \subseteq T_{MST}$.

DEF: e IS MINIMUM WEIGHT OUTGOING EDGE OF FRAGMENT T IF $e = \underset{u \in V_T,\, v \notin V_T}{\mathrm{ARGMIN}}\ \omega((u,v))$



$e = \mathrm{MWOE}(T)$



LEM II. T, T' ARE MST-FRAGMENTS OF G, THEN

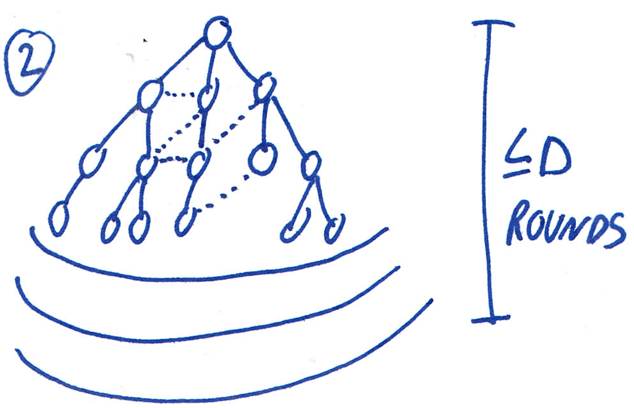
$T \cup \{\mathrm{MWOE}(T)\}$ (AND $T \cup T'$ (IF CONNECTED))

IS MST-FRAGMENT OF G.

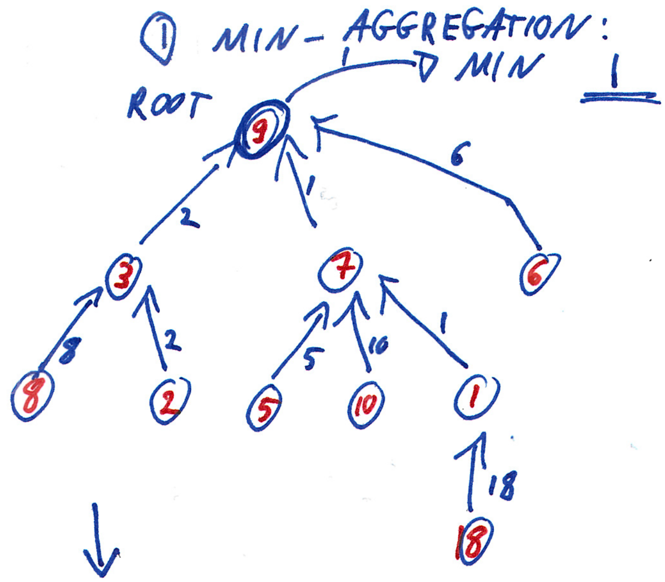
PROOF: BOOK / BASICS $\square$
+ USE UNIQUENESS

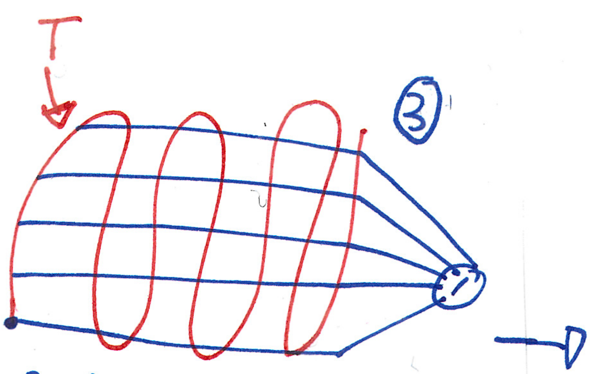
ALGO (PRIM): $T := \{1\}$; // INITIAL FRAGMENT
FOR $n-1$ ROUNDS DO
$T := T \cup \mathrm{MWOE}(T)$; ← BY MIN-AGGREGATION AND BROADCAST USING BREADTH-FIRST SEARCH TREE IN G. * (BFS-TREE)

ALGO BFS:

② 

① MIN-AGGREGATION:



③ 

BFS-TREE IN G HAS DIAM D

MIN-AGGREGATION TAKES TIME $O(D)$
(EVEN IF $\mathrm{DIAM}(T) \gg D$)

THM: ALGO (PRIM) TAKES TIME $O(n \cdot D)$ TO COMPUTE MST.

PROOF:
- COMPUTES MST: APPLIES LEM II $n-1$ TIMES.
- PERFORMS MIN-AGGREGATION $n-1$ TIMES $\Rightarrow$ TIME $O(n \cdot D)$. □

IDEA: GROW MANY FRAGMENTS IN PARALLEL.

ALGO (KRUSKAL): $T_1 := \{v_1\}, \dots, T_n := \{v_n\}$, $F = \{T_1, \dots, T_n\}$

FOR $\log n$ ROUNDS DO

IN PARALLEL, USE MIN-AGGREGATION IN T_i, NOT IN BFS-TREE OF G TO AVOID CONGESTION. → EACH $T_i \in F$ COMPUTES $e_i := MWOE(T_i) = (u, v)$

// ASSUME $u \in T_i$, $v \in T_j$, $j \neq i$

$\underbrace{T_{\min(i,j)}}_{\text{FRAGMENT ID}} := T_i \cup T_j$ // MERGE FRAGMENTS, BROADCAST FRAGMENT-ID TO ALL NODES IN THE NEW FRAG.

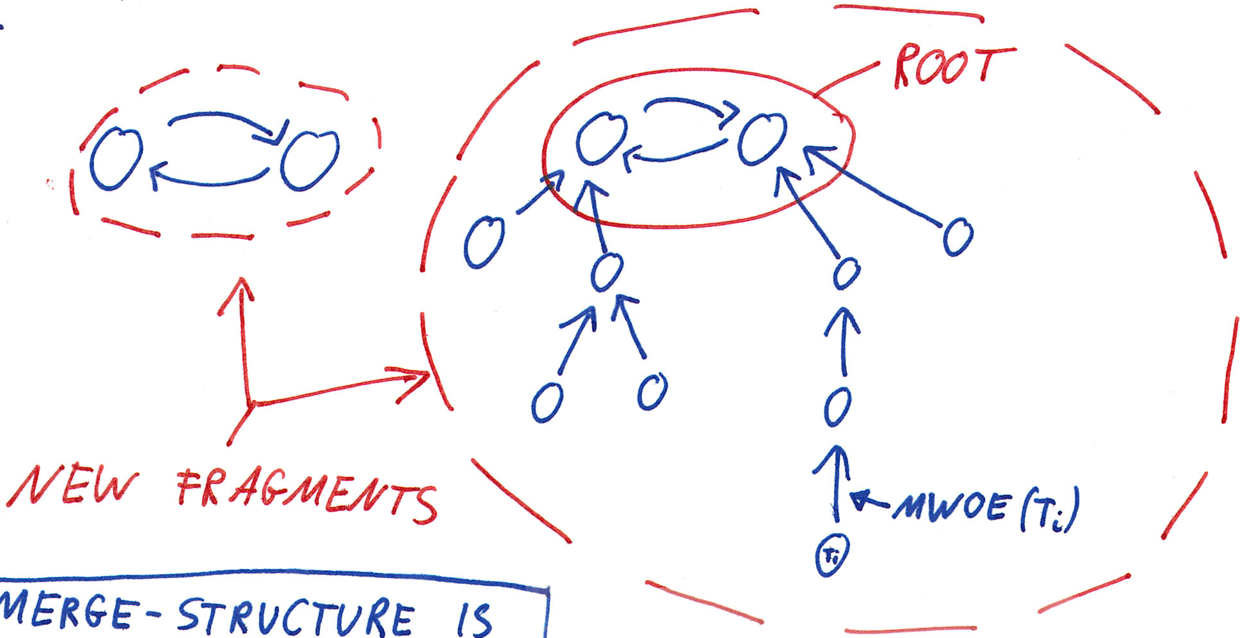
$T_{MST} := \{e_i \mid e_i$ WAS $MWOE(T_i)$ DURING THE ALGO$\}$;

LEM: ALGO (KRUSKAL) COMPUTES AN MST IN TIME $O(n \log n)$.

PROOF:
- MST: LEM II (OR BASIC LECTURE ON ALGORITHMS) + FRAGMENT SIZE DOUBLES (AT LEAST) IN EACH ITERATION. $\rightsquigarrow$ $\log n$ ITERATIONS SUFFICE
- EACH ITERATION PERFORMS A MIN-AGGREGATION ON EVERY FRAGMENT T_i (THIS TIME USING T_i FOR COMMUNICATION)

DIAM(T_i) MIGHT BE $\Theta(n)$. $\rightsquigarrow$ $O(n \log n)$ TIME □

FASTER? LOOK AT WHAT HAPPENS:



MERGE-STRUCTURE IS A DIRECTED TREE!

IDEA: COMPUTE MIS USING COLE VISHKIN ON THIS ~~UNDIRE~~ TREE.
NODES IN THE MIS INDUCE STARS OF CONSTANT DIAMETER.
EACH SUCH STAR BECOMES A NEW FRAGMENT.

ALGO (FAST-MST):

PHASE I

FOR I ROUNDS DO

$\hat{E} := \{ MWOE(T) \mid T \in F \}$; // $\hat{G} := (F, \hat{E})$; DIRECTED FOREST

$M :=$ MIS ON $\hat{G}$; // USE COLE-VISHKIN

EACH $T \in F \setminus M$ CHOOSES ONE (UNDIRECTED) $\hat{G}$-NEIGHBOR T' IN M;

T MERGES INTO T'; // ADAPTS T''s ID.

ANY $T \in M$ THAT DID NOT MERGE, MERGES INTO ONE NEIGHBOR

TAKE CARE OF LONELY FRAGMENTS TO GUARANTEE SIZE DOUBLES (AT LEAST)

PHASE II

FOR $\Theta(\log n)$ ROUNDS DO

NODE 1 • COLLECTS* $\hat{E} := \{ MWOE(T) \mid T \in F \}$

• COMPUTES NEW FRAGMENTS F'

• BROADCASTS THIS INFORMATION

*NOTE THAT ALTHOUGH FRAGMENTS MIGHT CHANGE, $\hat{E}$ DOES NOT.

FRAGMENTS IN F' MIGHT HAVE TOO LARGE DIAMETER TO COMPUTE MWOE(T).
AND IT IS $\hat{E}' \subseteq \hat{E}$. s.t. COMP. $\hat{E}$ SUFFICES.

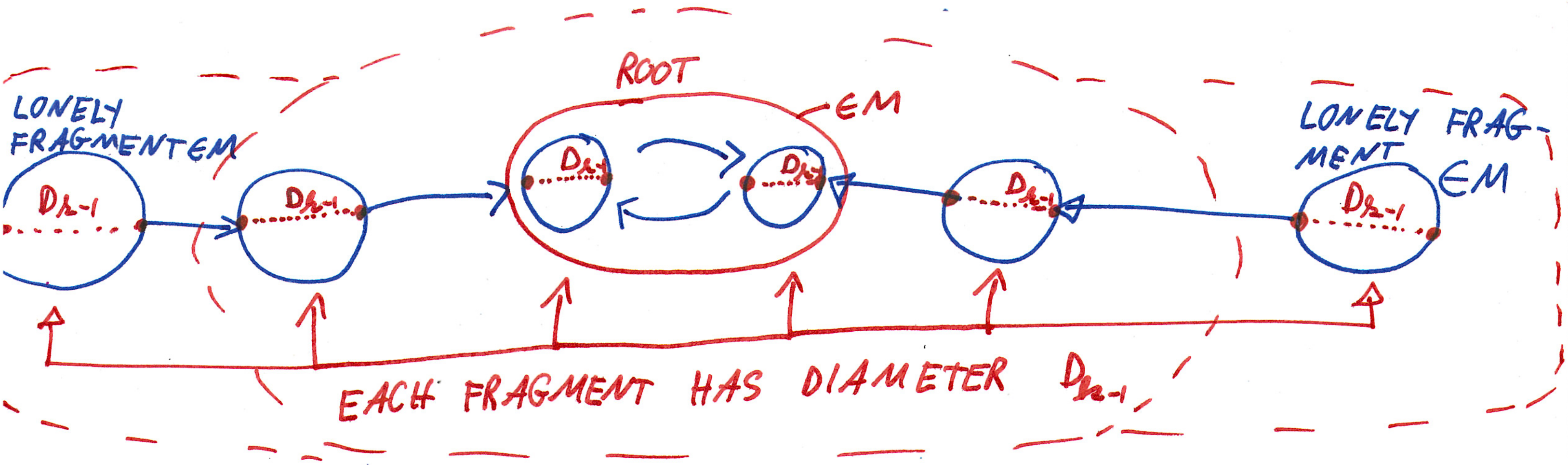
$T_{MST} := \{ e \mid e$ WAS MWOE(T) DURING THE EXECUTION ABOVE $\}$

DEF: (σ, ρ)- FRAGMENT: HAS • AT LEAST σ NODES
• DIAMETER AT MOST ρ.

LEM III: FRAGMENTS AFTER ROUND k IN PHASE I OF ALGO (FAST-MST) ARE $(2^k, 11^k)$- FRAGMENTS.

PROOF: INDUCTION

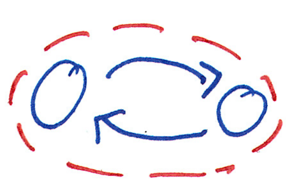
1) DIAMETER GROWS BY FACTOR ≤ 11 IN EACH ROUND



$\Rightarrow$ DIAMETER OF NEW FRAGMENT IS AT MOST $6D_{k-1}+5$ AS WE MERGE THE ROOT (2 FRAGMENTS) AND FRAGMENTS AT DISTANCE AT MOST 1 TO THE ROOT (IN CASE THE ROOT IS IN THE MIS) AS WELL AS LONELY FRAGMENTS ADJACENT TO THOSE

$\Rightarrow D_k \leq 11 D_{k-1} \quad (k \geq 2)$

2) SIZE OF FRAGMENTS GROWS BY FACTOR AT LEAST 2 IN EACH ROUND.



IN THE WORST CASE EACH FRAGMENT MERGES WITH ONLY ONE OTHER FRAGMENT

$\square$

THM: ALGO (FAST-MST) COMPUTES AN MST IN TIME $O(n^{0.775}\log n + D\log n)$ (FOR $I := \log_{22} n$).

PROOF:

- COMPUTES MST: ✓
  (ONLY ONE FRAGMENT LEFT)
- RUNTIME PHASE I: $\sum_{k=1}^{I} O(D_k \cdot \log^* n)$

  D_k: NEED D_k ROUNDS TO SIMULATE EACH ROUND OF COLE-VISHKIN

  $\log^* n$: COLE-VISHKIN MIS

  $$= \sum_{k=1}^{I} O(11^k \cdot \log^* n) = O(11^I \cdot \log^* n)$$

- RUNTIME OF PHASE II: $O((N + D_I + D)\log n)$

  D_I: TIME TO COMPUTE $\hat{E}$ EACH ITERATION

  $\log n$: EACH ITERATION SHRINKS F' BY A FACTOR OF AT LEAST TWO. AT THE BEGINNING $F' = F$.

  N: # FRAGMENTS AFTER PHASE I. $= |\hat{E}|$

  $N \leq n/2^I$, AS EACH FRAGMENT OF SIZE AT LEAST 2^I.

- TOTAL RUNTIME: $O(11^I \log^* n + \frac{n}{2^I}\log n + D\log n)$

  IS MINIMIZED FOR $I = \log_{22} n$. □

  $$11^I = 11^{\log_{22}(n)} = n^{\log 11/\log 22} = n^{0.775}$$

NOTE: A MORE CAREFUL ANALYSIS YIELDS $\tilde{O}(n^{0.613} + D)$.

FASTER?

OBSERVE: DIAMETER GROWS FASTER THAN SIZE
(UPPER BOUND) (LOWER BOUND)

IDEA: STOP FRAGMENTS FROM GETTING LARGE DIAMETER IN PHASE I.

DEF: FRAGMENT IS LONG IF IT HAS DIAMETER $\geq \sqrt{n}$.

NOTE: # LONG FRAGMENTS $\leq \sqrt{n}$

MODIFY ALGO (FAST-MST):
- IN PHASE I NO TWO LONG FRAGMENTS ARE ALLOWED TO MERGE. ~~with~~
- SET $I := \frac{\log n}{2} + 1$

LEM: AFTER PHASE I THERE ARE ONLY $(\sqrt{n}, \sqrt{n}\log n)$-FRAGMENTS.

PROOF: 1) SIZE OF FRAGMENTS IS $\geq \sqrt{n}$:
- ONCE A FRAGMENT IS LONG, IT HAS SIZE $\geq \sqrt{n}$.
- ASSUME A FRAGMENT NEVER BECOMES LONG IN PHASE I: LEM III SAYS IT HAS SIZE $\geq 2^{I} \geq \sqrt{n}$ AFTER $I = \frac{\log n}{2} + 1$ ROUNDS.

2) DIAMETER OF FRAGMENTS IS $\leq \sqrt{n} \log n$: ~~AFTER~~
- "SHORT" FRAGMENTS HAVE DIAM $\leq \sqrt{n}$
- ASSUME A FRAGMENT BECOMES LONG AT SOME POINT.
$\Rightarrow$ CAN ONLY MERGE WITH SHORT FRAGMENTS
$\Rightarrow$ DIAMETER GROWS BY AT MOST $11(\sqrt{n}-1)$ IN EACH ROUND.

$\Rightarrow$ DIAMETER GROWS BY AT MOST $\sqrt{n}\left(\frac{\log n}{2}+1\right)$.

AS A FRAGMENT HAS DIAMETER AT MOST $11(\sqrt{n}-1)$ AT THE TIME IT TURNS LONG (BY MERGING SHORT FRAGMENTS OF MAXIMAL DIAM $\sqrt{n}-1$), IT HAS DIAMETER AT MOST $O(\sqrt{n}\log n)$ AFTER PHASE I.

$\square$

THM: THE RUNTIME OF THE MODIFIED ALGO (FAST-MST) IS $O(\sqrt{n}\log^2 n \log^* n + D\log n)$.

PROOF: • PHASE I: $O(\underbrace{\sqrt{n}\log n}_{} \cdot \log^* n \cdot \log n)$

COLE-VISHKIN

DIAMETER OF FRAGMENTS $\leadsto$ TIME NEEDED TO SIMULATE EACH ROUND OF COLE-VISHKIN

$\log n$ MANY ROUNDS

• PHASE II: THERE ARE AT MOST $\frac{n}{\sqrt{n}}=\sqrt{n}$ MANY $(\sqrt{n},\sqrt{n}\log n)$- FRAGMENTS.

$\Rightarrow$ TIME $O((\sqrt{n}+\sqrt{n}\log n+D)\log n)$ TO COMPUTE AND COLLECT $\hat{E}$ EDGES $\log n$ TIMES.

• TOTAL : $O(\sqrt{n}\log^2 n\log^* n + D\log n)$ $\boxtimes$