6.S899 Distributed Graph Algorithms Lecture 3 MIT 09/27/2014

Material covered:

New and simple version of computing an MIS:
Chapter 2 of “Yves Metivier, John Michael Robson, Nasser Saheb-Djahromi and Akka Zemmari, An optimal bit complexity randomized distributed {MIS} algorithm, Distributed Computing volume 23, number 5-6, 2011, pages 331-340”. We follow Roger Wattenhofer’s lecture notes of a class on Principles of Distributed Computing at ETH Zurich, Chapter 7 “Maximal Independent Set”:
http://dcg.ethz.ch/lectures/podc_allstars/lecture/chapter7.pdf

Faster MIS for graphs of “small” degree:
Chapter 4 of “Leonid Barenboim, Michael Elkin, Seth Pettie and Johannes Schneider, The Locality of Distributed Symmetry Breaking,
53rd Annual IEEE Symposium on Foundations of Computer Science FOCS 2012, pages 321-330”

PREVIOUS LECTURES: COLORING, E.G. $\Delta+1$ COLORS IN TIME $O(\Delta)+\log^* n$

THIS LECTURE: MAXIMAL INDEPENDENT SET IN

+ANNOUNCE P-SET
+UPDATE GRADING

(A) $O(\log n)$ ROUNDS W.H.P.

(B) $O(\log \Delta \sqrt{\log n})$ ROUNDS DETERMINISTIC
if $\Delta \leq 2^{\sqrt{\log n}}$

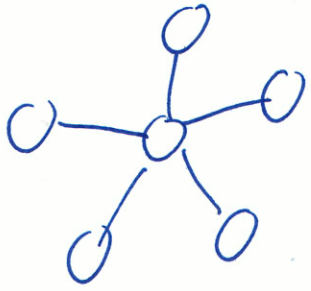
DEF: $S \subseteq V$ IS **INDEPENDENT** IN G, IF FOR ALL $u,u' \in S$ $(u,u') \notin E$.

MAXIMAL vs. **MAXIMUM?**

IF NO NODE IN $V \setminus S$ CAN BE ADDED TO S WITHOUT VIOLATING INDEPENDENCE

S IS IND. SET OF MAXIMUM CARDINALITY

QUESTION: MAXIMAL / MAXIMUM?



STAR:

1 OR $n-1$ NODES

$n-1$ NODES

NON-DISTRIB.

GREEDY $O(n+m)$

↓

CONSIDER DISTRIBUTED VERSION

NP-HARD EVEN $n^{1/2-\varepsilon}$ APPROX IS HARD

OBSERVE: • COMPUTING C-COLORING IN T ROUNDS IMPLIES MIS IN TIME $C+T$

(I) $O(\Delta) + \log^* n$ $(C = \Delta+1)$

(II) $\log^*(n) + O(1)$ IF $\Delta = \Theta(1)$ $(C = \Delta+1)$

• COMPUTING MIS IN T ROUNDS IMPLIES $\Delta+1$-COLORING IN TIME $T \cdot (\Delta+1)$

$\rightsquigarrow$ (II) IS TIGHT DUE TO COLORING LOWER BOUND (UP TO CONSTANTS)

QUESTION: CAN WE DO FASTER? HOW?

ALGO: $M := \emptyset$

FOR $i = 0, \dots, \Theta(\log n)$ DO

PHASE:
- $A_i := V \setminus \Gamma(M)$ // ACTIVE NODES
- EACH $v \in A_i$ • CHOOSES RANDOM $r(v) \in [0,1] \subseteq \mathbb{R}$
  - SENDS $r(v)$ TO NEIGHBORS
  - JOINS M IF $r(v) > r(w)$ FOR ALL $w \in \Gamma_i(v)$

$\Gamma_i(v)$: NEIGHBORS IN $G|_{A_i} = (A_i, E_i)$

RETURN M

THM I: COMPUTES MIS IN $\Theta(\log n)$ ROUNDS w.h.p.

PROOF: COMPUTES IS: ✓ AS $A_i := V \setminus T(M)$ AND $\tau(v) < \tau(w)$

MAXIMAL IS: IF $A_{\Theta(\log n)} = \emptyset$ [w.h.p.

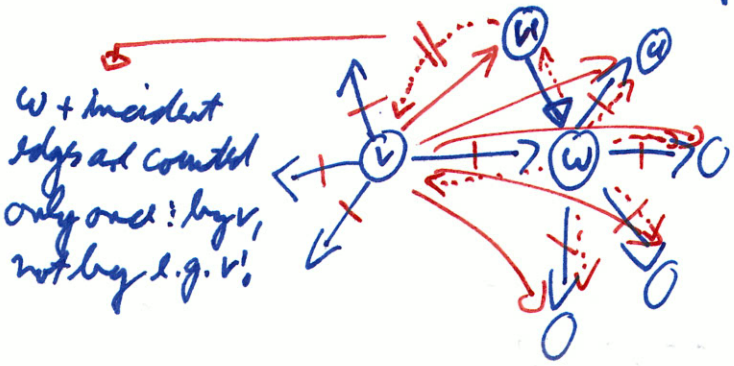
DEFINE: RAND. VAR. $X_i := |E_{i+1}| - |E_i|$

LEMMA: $\mathbb{E}[X_i] \geq |E_i|/2$

PROOF: UNDIRECTED $\{a, v\} \in E_i$

DIRECTED (a,v) (v,u)

EVENT $(v \to w)_i$: $v \in A_i$ JOINS M IN PHASE i, $(v,w) \in E_i$, $\tau(v) \geq \tau(u)$ FOR ALL $u \in T_i(w)$



$$\Pr[(v \to w)_i] > \frac{1}{\overrightarrow{deg}_i(v) + \overrightarrow{deg}_i(w)}$$

$\vec{T}_i(w)$ EDGES REMOVED

$$\text{RV } X_{(v \to w)_i} := \begin{cases} \overrightarrow{deg}_i(w) & : (v \to w)_i \text{ OCCURS} \\ 0 & : \text{OTHERWISE} \end{cases}$$

LINEARITY OF EXPECTATION

$X_i \leq \sum_{\{u,v\} \in E_i} X_{(u \to v)_i}$

$$\mathbb{E}[2X_i] \geq \sum_{\{v,w\} \in E} \mathbb{E}[X_{(v \to w)_i}] + \mathbb{E}[X_{(w \to v)_i}]$$

DIRECTED

$$= \sum \Pr[(v \to w)_i] \cdot \overrightarrow{deg}_i(w) + \Pr[(w \to v)_i] \cdot \overrightarrow{deg}_i(v)$$

$$> \sum \frac{\overrightarrow{deg}_i(w) + \overrightarrow{deg}_i(v)}{\overrightarrow{deg}_i(w) + \overrightarrow{deg}_i(w)} = |E|$$

$$\Rightarrow \mathbb{E}[X_i] > |E|/2$$

DEF: RV $Y_i := \begin{cases} 1 : X_i \geq |E_i|/4 \\ 0 : \text{——} \end{cases}$, $Y := \sum_{i=0}^{\Theta(\log n)} Y_i$

INDEPENDENT RVs

CLAIM: $PR[Y_i] \geq 1/3$

PROOF: ASSUME NOT $\Rightarrow \mathbb{E}[X] < \frac{2}{3} \cdot \frac{|E_i|}{4} + \frac{1}{3} \cdot |E_i|$

$= \frac{|E_i|}{2}$ ↯ LEM II □

VERSION OF CHERNOFF: $|\mathbb{E}[Y] - Y| \leq C(\log n + \sqrt{\mathbb{E}[Y] \log n})$ W.H.P.

PROOF OF THM I: $\mathbb{E}[Y] = \Theta(\log n) \cdot \frac{1}{3}$

$\overset{\text{W.H.P.}}{\Rightarrow} Y > \Theta(\log n) - C(\log n + \sqrt{\Theta(\log^2 n)})$

$= \Theta(\log n)$

↑ CONST LARGE ENOUGH

$\overset{\text{W.H.P}}{\Rightarrow} \Theta(\log n)$ PHASES REMOVE ALL EDGES

REMAINING SINGLETON NODES JOIN ☒ WHP

NOTE: CAN USE $\Theta(\log n)$ BIT TO ENCODE $r(v)$.

MORE ANALYSIS: REDUCE MESSAGE-COMPLEXITY TO $\Theta(1)$ BITS

WITH $\Theta(\log n)$ ROUNDS W.H.P.

LOWER BOUND: $\Omega(\sqrt{\log n})$.

THM: MIS, $O(\log \Delta \sqrt{\log n})$.

↳ CAN ASSUME $\Delta \leq 2^{\sqrt{\log n}}$

ALG: $O(\log \Delta)$ TIMES DO

// REDUCE DEGREE BY FACTOR ≥ 2

$I_0 := \emptyset$

PHASE I

FOR $i = 0, \ldots, K$ DO (K: $c_i \log n$)

$A_i := V \setminus \Gamma(I_i)$ // ACTIVE NODES

$M_i' := \emptyset$

EACH $v \in A_i$ JOINS M_i' WITH PROB $\frac{1}{\Delta+1}$

$I_{i+1} := I_i \cup \{v \in M_i' \mid \Gamma_i(v) \cap M_i' = \{v\}\}$

↑ NEIGHBORS IN $G|_{A_i}$

// NEEDS 1-BIT MSG

// NICE STRUCTERE REMAINS

PHASE II

$U := \{v \in A_K \mid \deg_K(v) \geq \Delta/2\}$

↑ DEG. IN $G|_{A_i}$

EACH $v \in V$ COMPUTES MIS $I(C)$ FOR EACH CONNECTED COMPONENT C OF $G|_U$ THAT IS CONTAINED IN v'S $5\sqrt{\log n}$-NEIGHBORHOOD

RETURN $M := I_K \cup \bigcup_{C:\text{ CONNECTED IN } G|_U} I(C)$

LEM: $PR[u \in A_{i+1} \mid deg_i(u) \geq \Delta/2] \leq 1-(1-e^{-1/2})e^{-1} =: p_0 = \Theta(1)$

PROOF:

$$p_1 := PR[\exists\, v_j \in \Gamma_i(u) \text{ s.t. } v_j \in M_i'] \geq 1-\left(1-\frac{1}{\Delta+1}\right)^{\Delta+1} \geq 1-e^{-\frac{\Delta/2+1}{\Delta+1}} \geq 1-e^{-1/2}$$

↑ FOR SIMPLICITY WE IGNORE DEPENDENCIES FOR NOW!

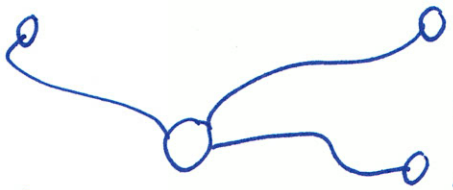
↓

$$p_2 := PR[v_j \in M' \text{ joins } I_{i+1}] \geq \left(1-\frac{1}{\Delta+1}\right)^{|\Gamma_i(v_j)|} > e^{-1}$$

$p_0 = 1 - p_1 \cdot p_2$ ☒

QUESTION: HOW DOES THIS HELP?

WHAT IF PATHS LONG?



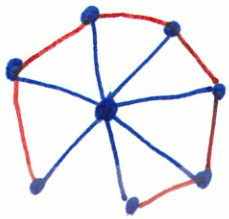
EACH PATH MIGHT STOP!

SHOW COMPONENTS HAVE SMALL WEAK DIAMETER

DEF: WEAK DIAMETER (W.D.): $V' \subseteq V$, $\max_{u,v \in V'} d_G(u,v)$

QUESTION: HOW CAN WE PROVE THIS?

76 NODES!



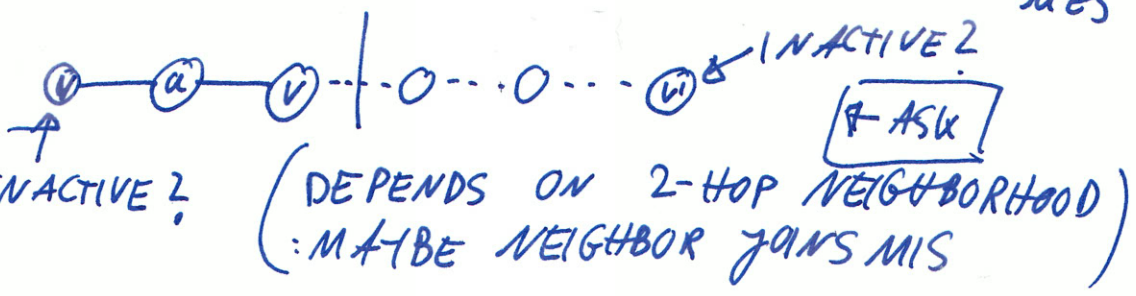
—— HAS DIAM 6

WEAK DIAM 2 IN ≈≈

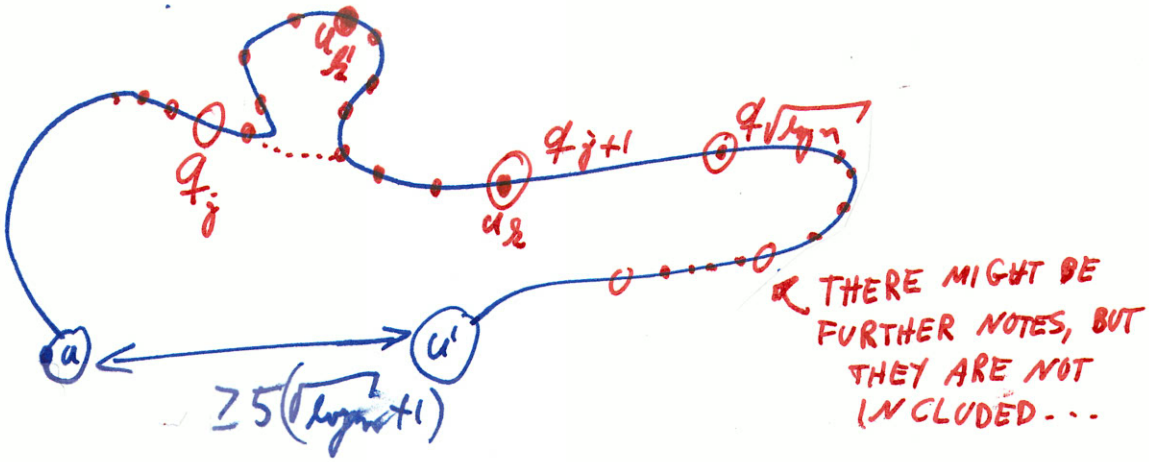
DEF: $\mathcal{P}$:= SET OF ALL PATHS BETWEEN ANY NODES OF DISTANCE $\geq 5(\sqrt{\log n}-1)$

$P = (u_1, \ldots, u_r) \in \mathcal{P}$ WE SHOW: AT LEAST ONE $u_i \notin U$! (*)

~~CLAIM~~ NOTE: TO GET INDEPENDENCE OF EVENTS $\{u \in A_{i+1}\}_{u \in S}$



CAN ONLY CONSIDER NODES AT DISTANCE 5



$Q(P) = (q_1, \dots, q_{\sqrt{\log n}}) \subset V(P)$ s.t. $d(u,u') \geq 5 \; \forall u, u' \in Q(P)$.

$q_1 := u_1$ ← SET OF NODES

ASSUME q_j KNOWN, $q_{j+1} := u_r$

← MAX INDEX S.T. $d(q_j, u_r) = 5$

$(\Rightarrow d(q_{j'}, q_{j+1}) > 5 \; \forall j' < j)$

$\Rightarrow d(q_j, u_r) \geq 5(\sqrt{\log n} - j)$

$Q := \{Q(P) \mid P \in \mathcal{P}\}$ $\quad |Q| \le n \cdot \underbrace{\Delta^{5(\sqrt{\log n})}}$ DIST 5 BETWEEN $q_{j_i}, q_{j_{i+1}}$

LEM: COMPONENTS IN U HAVE W.D. $\le 5\sqrt{\log n}$ WHP.

PROOF: BOUND $P_3 := PR[\exists P \in \mathcal{P} \text{ s.t. } V(P) \subseteq U]$

$\uparrow$ deg $\ge \Delta/2$ ELSE NOT $\subseteq U$

FIX PATH $P \in \mathcal{P}$, $\deg(v) \ge \Delta/2$ FOR ALL $v \in V(P)$

$\mathcal{E}(n) := PR[Q(P) \subseteq U] \le p_0^{-|Q(P)| \cdot K}$

$K \leftarrow \sqrt{\log n}$, $|Q(P)| \leftarrow c_2\sqrt{\log n}$

$= p_0^{-c_2 \log n}$

$P_3 :\le \mathcal{E}(n) \cdot |Q| = p_0^{-c_2 \log n} \cdot n\Delta^{5\sqrt{\log n}}$ // $\Delta \le 2^{\sqrt{\log n}}$

$= 1/\text{poly}\, n$ FOR SOME c_2.

$\Rightarrow$ WHP W.D. $\le 5\sqrt{\log n}$ ☒

PROOF OF MAIN THM:

$O(\log \Delta)$ TIMES:

- PHASE I : $O(\sqrt{\log n})$ COMPUTES IS
- PHASE II : $O(\sqrt{\log n})$, AS WEAK DIAM $\le 5(\sqrt{\log n - 1})$ (COLLECT $5\sqrt{\log n - 1}$-neighborhood and send back).

MAKES IS MAXIMAL W.R.T. NODES OF DEG $\ge \Delta/2$

$\hookrightarrow$ MIS ON G ☒ ☒