DEFN.

A MATROID IS AN ORDERED PAIR $M = (S, \mathcal{I})$ SATISFYING.

1.) S IS A FINITE NON-EMPTY SET, AND $\mathcal{I} \subseteq \mathcal{P}(S)$. THE MEMBERS OF $\mathcal{I}$ ARE CALLED THE INDEPENDENT SUBSETS OF S

2.) HEREDITARY PROPERTY
IF $B \in \mathcal{I}$ AND $A \subseteq B$, THEN $A \in \mathcal{I}$.

3.) EXCHANGE PROPERTY
IF $A \in \mathcal{I}$, $B \in \mathcal{I}$, AND $|A| < |B|$, THEN THERE EXISTS $x \in B - A$ SUCH THAT $A \cup \{x\} \in \mathcal{I}$.

NOTE THAT (2) IMPLIES THAT $\emptyset \in \mathcal{I}$ (PROVIDED $\mathcal{I}$ IS ITSELF NON-EMPTY.)

EX. (MATRIC MATROIDS)
LET P BE A (RECTANGULAR) MATRIX AND LET $S = \{\text{rows of } P\}$, CONSIDERED AS VECTORS.
LET

$$\mathcal{I} = \{A \subseteq S \mid A \text{ is linearly independent}\}$$

Obviously S is finite and non-empty, and $\mathcal{I} \subseteq \mathcal{P}(S)$. Properties (2) & (3) are elementary facts of linear algebra.

Similarly we could let S be the columns of D!

Ex. (Graphic matroids)
Let $G=(V,E)$ be an undirected graph. Let $S=E$ and

$$\mathcal{I}=\{A \subseteq S \mid \text{subgraph } (V,A) \text{ is acyclic}\}.$$

(1) is clearly satisfied. (2) holds since any subset of an acyclic set of edges is acyclic. (By removing edges we cannot create cycles.)

We prove the exchange property (3):

Let $A, B \in \mathcal{I}$ and suppose $|A|<|B|$. Then the forest (V,A) contains exactly $|V|-|A|$ trees and (V,B) contains $|V|-|B|$ trees.

(Pf: suppose (V,A) contains m trees:
$T_i=(V_i,E_i)$, $1 \le i \le m$. Then $|E_i|=|V_i|-1$

AND

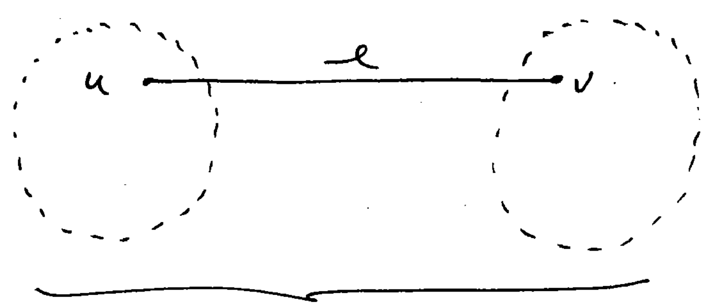
$$|A| = \sum_{i=1}^{m} |E_i| = \sum_{i=1}^{m} (|V_i| - 1) = |V| - m,$$

$\therefore m = |V| - |A|$. LIKEWISE (V,B) CONTAINS $|V| - |B|$ TREES. //.]

NOW $|A| < |B|$ IMPLIES $|V| - |B| < |V| - |A|$, SO (V,B) CONTAINS FEWER TREES THAN DOES (V,A).

THEREFORE (V,B) CONTAINS A TREE T WHICH HAS VERTICES IN TWO DISTINCT TREES OF FOREST (V,A). (OTHERWISE EACH TREE OF (V,B) IS A SUBTREE OF ONE IN (V,A), WHICH IMPLIES (V,B) HAS AT LEAST AS MANY TREES AS FOREST (V,A). $\times$)

THUS T, BEING CONNECTED, CONTAINS AN EDGE $e = uv$ WHOSE ENDS LIE IN DIFFERENT TREES OF FOREST (V,A)



$e \in E(T) \subseteq B$

DISTINCT TREES OF (V,A)

THUS BY ADDING e TO A NO CYCLE IS CREATED (INSTEAD TWO TREES ARE JOINED)

$$\therefore \ (V, A \cup \{e\}) \text{ is acyclic}$$

$$\therefore \ A \cup \{e\} \in \mathcal{I},$$

AND THE EXCHANGE PROPERTY HOLDS.

DEFN

GIVEN A MATROID $M = (S, \mathcal{I})$ AND $A \in \mathcal{I}$, WE CALL $x \in S - A$ AN EXTENSION OF A IF $A \cup \{x\} \in \mathcal{I}$.

$A \in \mathcal{I}$ IS CALLED MAXIMAL IF IT HAS NO EXTENSIONS, i.e.

$$\forall x \in S - A : \ A \cup \{x\} \notin \mathcal{I}.$$

A MAXIMAL INDEPENDENT SET IS ALSO CALLED A BASE OF THE MATROID.

IN A MATRIC MATROID A BASE IS JUST A VECTOR SPACE BASIS FOR THE ROW SPACE OF THE UNDERLYING MATRIX.

IN A GRAPHIC MATROID A BASE FORMS THE EDGE SET OF A SPANNING TREE.

DEFN:
A SUBSET $D \subseteq S$ IS CALLED DEPENDENT IF IT IS NOT INDEPENDENT, i.e. $D \notin \mathbb{I}$

A SUBSET $C \subseteq S$ IS CALLED A cycle IF IT IS MINIMAL WITH RESPECT TO THE PROPERTY OF BEING DEPENDENT. i.e. C IS DEPENDENT WHILE EACH OF ITS SUBSETS ARE INDEPENDENT.

OBVIOUSLY A cycle IN A GRAPHIC MATROID IS A cycle IN THE ORDINARY SENSE.

THEOREM
ANY TWO BASES IN A MATROID HAVE THE SAME CARDINALITY.

THIS COMMON CARDINALITY IS CALLED THE RANK OF THE MATROID, DENOTED rank(M).

PROOF:

LET $A, B \in \mathcal{I}$ BE BASES AND SUPPOSE $|A| < |B|$. BY THE EXCHANGE PROPERTY THERE EXISTS $x \in B - A$ SUCH THAT $A \cup \{x\} \in \mathcal{I}$, CONTRADICTING THAT A IS MAXIMAL. $|B| < |A|$ LEADS TO A SIMILAR CONTRADICTION. THEREFORE $|A| = |B|$ AS CLAIMED.

DEFN.

A WEIGHTED MATROID IS A MATROID $M = (S, \mathcal{I})$ EQUIPPED WITH A (STRICTLY POSITIVE) WEIGHT FUNCTION ON ITS UNDERLYING SET S.

$$w : S \longrightarrow \mathbb{R}^+$$

WE EXTEND THIS WEIGHT FUNCTION TO SUBSETS $A \subseteq S$ BY SUMMATION.

$$w(A) = \sum_{x \in A} w(x)$$

A SET $A \in \mathcal{I}$ IS CALLED OPTIMAL IF IT HAS MAXIMUM WEIGHT AMONG ALL INDEPENDENT SUBSETS OF S.

NOTE THAT AN OPTIMAL SET A IS NECESSARILY MAXIMAL SINCE OTHERWISE WE COULD FIND AN $x \in S - A$ SUCH THAT $A \cup \{x\} \in \mathcal{I}$, AND THEN

$$w(A \cup \{x\}) = w(A) + w(x) > w(A)$$

(USING HERE THAT $w(x) > 0$), WHICH CONTRADICTS THAT A IS OPTIMAL.

<u>PROBLEM</u>

GIVEN A WEIGHTED MATROID $M = (S, \mathcal{I})$, $w: S \to \mathbb{R}^+$, FIND AN OPTIMAL SUBSET OF S. i.e. FIND $A \subseteq S$ SUCH THAT

- $A \in \mathcal{I}$
- $w(A) \geq w(B)$ FOR ALL $B \in \mathcal{I}$.

MANY OPTIMIZATION PROBLEMS FOR WHICH A GREEDY STRATEGY IS APPLICABLE CAN BE SHOWN, BY REFORMULATION, TO BE EQUIVALENT TO THE ABOVE PROBLEM.

MORE IMPORTANTLY, ANY PROBLEM WHICH CAN BE REFORMULATED AS ABOVE CAN BE SOLVED BY A GREEDY STRATEGY.

Ex

FIND A MINIMUM WEIGHT SPANNING TREE IN A WEIGHTED GRAPH $G = (V, E, w)$.

CONSIDER THE GRAPHIC MATROID M_G OF G WITH WEIGHT FUNCTION

$$w'(e) = w_0 - w(e)$$

WHERE w_0 IS A FIXED NUMBER GREATER THAN ALL EDGE WEIGHTS IN G. THUS $w'(e) > 0$ FOR ALL $e \in E$.

RECALL THAT A MAXIMAL INDEPENDENT SET (BASE) $A \subseteq E$ CORRESPONDS TO A SPANNING TREE IN G. OBSERVE

$$\begin{aligned} w'(A) &= \sum_{e \in A} w'(e) \\ &= \sum_{e \in A} (w_0 - w(e)) \\ &= |A| \cdot w_0 - w(A) \\ &= (|V|-1) w_0 - w(A). \end{aligned}$$

THUS ANY SPANNING TREE WHICH MAXIMIZES w' NECESSARILY MINIMIZES w, AND CONVERSELEY.

THE FOLLOWING ALGORITHM (SOMETIMES CALLED "THE GREEDY ALGORITHM") SOLVES THE OPTIMAL SUBSET PROBLEM IN A WEIGHTED MATROID.

GREEDY(M, w)
1.) $A \leftarrow \emptyset$
2.) SORT $S[M]$ IN DESCENDING ORDER BY WEIGHT
3.) for each $x \in S[M]$ // TAKEN IN ORDER
4.) if $A \cup \{x\} \in \mathcal{I}[M]$
5.) $A \leftarrow A \cup \{x\}$
6.) return A

THEOREM

GREEDY(M, w) RETURNS AN INDEPENDENT SET OF MAXIMUM WEIGHT IN M.

PROOF

OBSERVE THAT $A \in \mathcal{I}$ AT EACH STAGE OF EXECUTION SO THE SET RETURNED IS CERTAINLY INDEPENDENT. IT IS ALSO MAXIMAL BY ITS VERY CONSTRUCTION.

LET B BE ANY MAXIMAL INDEPENDENT SET OF M. THEN $|A| = |B| = r = \text{rank}(M)$. WE MUST SHOW THAT $w(A) \geq w(B)$, WHENCE A IS OPTIMAL.

WRITE $A = \{x_1, \ldots, x_r\}$ AND $B = \{y_1, \ldots, y_r\}$ WHERE THE ELEMENTS ARE INDEXED BY DECREASING WEIGHTS. IN PARTICULAR, THE ELEMENTS OF A ARE LISTED IN THE ORDER SELECTED BY GREEDY(M,w).

IF $A = B$ THEN $w(A) = w(B)$ AND THERE IS NOTHING TO PROVE, SO ASSUME $A \neq B$.

LET x_k BE THE FIRST ELEMENT OF A NOT IN B, i.e.

- $x_i = y_i$ FOR $1 \leq i < k$
- $x_k \neq y_k$

LET $A' = A - \{x_k\}$. THEN $A' \in \mathcal{I}$ AND $|A'| = |A| - 1 < |B|$. BY THE EXCHANGE PROPERTY THERE EXISTS $y \in B - A'$ SUCH THAT $A' \cup \{y\} \in \mathcal{I}$.

<u>CLAIM</u>: $w(x_k) \geq w(y)$

<u>PROOF</u>: OBSERVE $\{x_1, \ldots, x_{k-1}, y\} \subseteq A' \cup \{y\}$ SO THAT $\{x_1, \ldots, x_{k-1}, y\} \in \mathcal{I}$. IF $w(x_k) < w(y)$ THEN GREEDY(M,w) WOULD HAVE CHOSEN y ON THE k^{TH} ITERATION OF LOOP 3–5 RATHER THAN x_k. $\therefore w(x_k) \geq w(y)$.

LET $A_1 = A' \cup \{y\} = (A - \{x_k\}) \cup \{y\}$.
THEN $A_1 \in \mathcal{I}$ AND THE ABOVE CLAIM SHOWS $w(A) \geq w(A_1)$. ALSO NOTE A_1 HAS ONE MORE ELEMENT IN COMMON WITH B THAN A DOES, NAMELY y.

IF $A_1 = B$ THEN $w(A) \geq w(B)$ AND WE ARE DONE. OTHERWISE WE REPEAT THIS PROCESS WITH A_1 IN PLACE OF A TO OBTAIN $A_2 \in \mathcal{I}$ WHERE $w(A_1) \geq w(A_2)$ AND A_2 HAS MORE IN COMMON WITH B THAN A_1.

CONTINUING IN THIS FASHION WE CONSTRUCT A SEQUENCE OF INDEPENDENT SETS STARTING AT A AND ENDING AT B WITH DECREASING WEIGHTS:

$$w(A) \geq w(A_1) \geq w(A_2) \geq \dots \geq w(B)$$

SHOWING THAT $w(A) \geq w(B)$ AS REQUIRED. ///

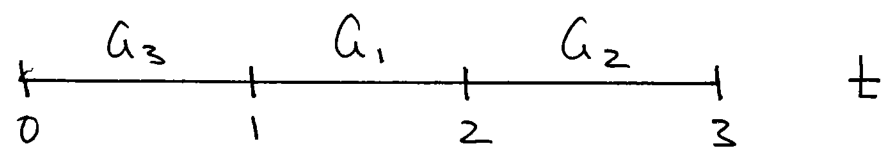
<u>RUN TIME</u>: LET $n = |S|$. IF THE TEST ON (4) TAKES $\Theta(f(n))$ TIME AND THE SORT ON (2) TAKES $\Theta(n \lg n)$, THEN GREEDY RUNS IN TIME $\Theta(n \lg n + n f(n))$.

Scheduling Unit Time Tasks (16.5)

A <u>unit time task</u> is simply a job which requires one unit of time to complete.

Given a set $S = \{a_1, a_2, \ldots, a_n\}$ of n unit time tasks, a <u>schedule</u> for S is simply a permutation of S giving the order in which the tasks are to be performed. We assume that any schedule begins at time $t = 0$ and ends at time $t = n$.

<u>Ex.</u> $S = \{a_1, a_2, a_3\}$ schedule: $a_3\, a_1\, a_2$



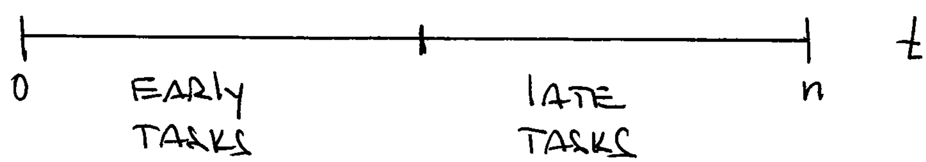
Suppose each task $a_i \in S$ has a deadline d_i satisfying $1 \le d_i \le n$, and a penalty $w_i \ge 0$ to be paid if a_i finishes later than its deadline $(1 \le i \le n)$.

<u>Problem</u>

Determine a schedule for S which minimizes the total penalty for missed deadlines.

Consider any schedule for S. We say a task is <u>early</u> in this schedule if it finishes before its deadline. Otherwise the task is said to be <u>late</u> in the schedule.

A schedule is said to be in <u>early-first form</u> if all early tasks are completed before any late tasks are started.



If some late task a_j is performed before some early task a_i, then upon swapping a_i with a_j, a_j is still late and a_i is still early. Thus any schedule can be placed in early-first form without changing it's penalty.

A schedule is said to be in <u>cannonical form</u> if early tasks precede late tasks, and the early tasks are scheduled in order of increasing deadlines.

ONE CAN GO FROM EARLY-FIRST FORM TO CANNONICAL FORM WITHOUT CHANGING THE PENALTY OF A SCHEDULE.

PROOF

SUPPOSE a_i AND a_j ARE EARLY TASKS WHICH FINISH AT TIMES $t_i \le d_i$ AND $t_j \le d_j$ RESPECTIVELY, AND SUPPOSE ALSO THAT $t_j < t_i$ AND $d_i < d_j$.

UPON SWAPPING a_i WITH a_j IN THIS SCHEDULE WE SEE THAT a_j FINISHES AT TIME

$$t_j' = t_i \le d_i < d_j$$

AND a_i FINISHES AT TIME

$$t_i' = t_j < t_i \le d_i .$$

HENCE BOTH TASKS ARE EARLY IN THE NEW SCHEDULE, AND THE PENALTY FOR THE NEW SCHEDULE IS THE SAME AS FOR THE OLD.

CLEARLY ANY EARLY-FIRST SCHEDULE CAN BE PLACED IN CANNONICAL FORM BY PERFORMING SWAPS OF THIS KIND. ///

Ex. $n = 7$

TASK:	a_1	a_2	a_3	a_4	a_5	a_6	a_7
d_i:	4	1	5	2	6	1	6
w_i:	1	2	1	1	2	3	1

A SCHEDULE WITH PENALTY 6:

t	0–1	1–2	2–3	3–4	4–5	5–6	6–7
Task	a_4	a_6	a_2	a_3	a_5	a_7	a_1
		LATE	LATE				LATE

EARLY-FIRST FORM:

t	0–1	1–2	2–3	3–4	4–5	5–6	6–7
Task	a_4	a_7	a_5	a_3	a_2	a_6	a_1
	EARLY				LATE		

CANONICAL FORM:

t	0–1	1–2	2–3	3–4	4–5	5–6	6–7
Task	a_4	a_3	a_5	a_7	a_2	a_6	a_1
	EARLY				LATE		

AN OPTIMAL SCHEDULE (PENALTY = 2):

t	0–1	1–2	2–3	3–4	4–5	5–6	6–7
Task	a_6	a_4	a_1	a_3	a_7	a_5	a_2
	EARLY						LATE

DEFINE

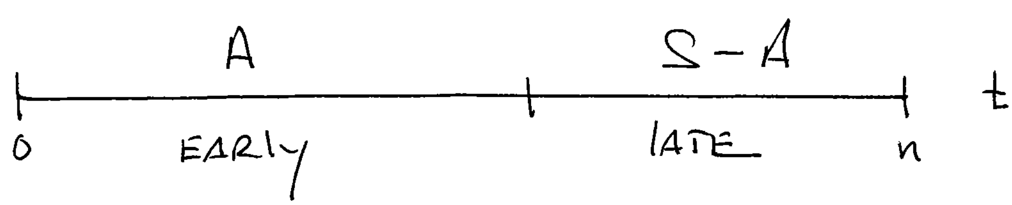
$$\mathcal{I} = \{ A \subseteq S \mid \text{THERE EXISTS A SCHEDULE FOR } S \text{ IN WHICH ALL TASKS IN } A \text{ ARE EARLY} \}.$$

OBVIOUSLY THE SET OF EARLY TASKS IN SOME SCHEDULE CONSTITUTES AN INDEPENDENT SET.

THEOREM

$M = (S, \mathcal{I})$ IS A MATROID.

OBSERVE THAT THE PROBLEM OF FINDING A SCHEDULE FOR S WHICH MINIMIZES THE PENALTY FOR MISSED DEADLINES IS EQUIVALENT TO FINDING A MAXIMUM WEIGHT INDEPENDENT SET IN THIS MATROID.



i.e. WE CAN MINIMIZE THE FINES WHICH MUST BE PAID BY MAXIMIZING THE FINES WHICH NEED NOT BE PAID.

DEFINE FOR $t = 0, 1, \ldots, n$ AND $A \subseteq S$:

$$N_t(A) = (\text{\# OF TASKS } a_i \in A \text{ SUCH THAT } d_i \leq t)$$

NOTE THAT $N_0(A) = 0$ AND $N_n(A) = |A|$ FOR ANY $A \subseteq S$.

LEMMA

LET $A \subseteq S$. THE FOLLOWING ARE EQUIVALENT.

(1) $A \in \mathcal{I}$

(2) $N_t(A) \leq t$ FOR $t = 0, 1, \ldots, n$

(3) IF THE TASKS OF A ARE PERFORMED IN ORDER OF INCREASING DEADLINES (STARTING AT $t=0$ AND WITH NO IDLE TIME BETWEEN TASKS), THEN NO TASK IS LATE.

PROOF:

(1) $\Rightarrow$ (2)

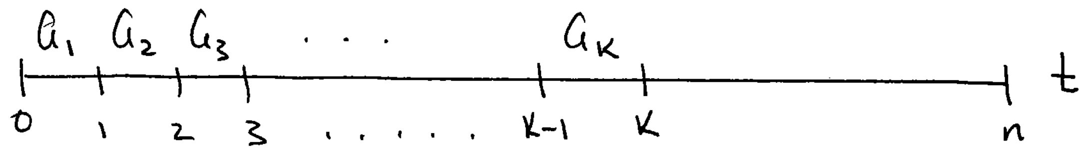
EQUIVALENTLY WE PROVE THE CONTRAPOSITIVE: NOT (2) $\Rightarrow$ NOT (1). SUPPOSE $N_t(A) > t$ FOR SOME t. THEN THERE ARE MORE THAN t TASKS IN A WHICH MUST FINISH BEFORE TIME t. THESE TASKS CANNOT BE SCHEDULED WITHOUT AT LEAST ONE OF THEM BEING LATE. $\therefore A \notin \mathcal{I}$.

(2) $\Rightarrow$ (3)

ASSUME (2) HOLDS AND SUPPOSE THE TASKS IN A ARE SCHEDULED BY INCREASING DEADLINES. BY RE-INDEXING THE ELEMENTS OF S IF NECESSARY, WE MAY ASSUME

$$A = \{a_1, a_2, \ldots, a_k\} \subseteq S$$

WITH DEADLINES $d_1 < d_2 < \cdots < d_k$. OUR (PARTIAL) SCHEDULE IS THEN



SINCE $d_1 \geq 1$, TASK a_1 IS NOT LATE. BUT ALSO

$N_1(A) \leq 1 \Rightarrow d_2 \geq 2 \Rightarrow a_2$ NOT LATE
$N_2(A) \leq 2 \Rightarrow d_3 \geq 3 \Rightarrow a_3$ NOT LATE
$N_3(A) \leq 3 \Rightarrow d_4 \geq 4 \Rightarrow a_4$ NOT LATE
$\vdots$
$N_{k-1}(A) \leq k-1 \Rightarrow d_k \geq k \Rightarrow a_k$ NOT LATE

$\therefore$ NO TASK IN A IS LATE.

(3) $\Rightarrow$ (1) IS OBVIOUS. ///

EXERCISE

WRITE AN ALGORITHM WHICH DETERMINES WHETHER OR NOT $A \subseteq S$ IS INDEPENDENT. (HINT: USE PART (2) OF THE PRECEDING LEMMA, AND RECALL COUNTING SORT.)

EXERCISE

WRITE AN ALGORITHM WHICH DETERMINES A SCHEDULE OF UNIT TIME TASKS WITH MINIMUM TOTAL PENALTY. (HINT: BASE YOUR ALGORITHM ON THE GREEDY ALGORITHM FOR WEIGHTED MATROIDS.)

IT REMAINS ONLY TO PROVE THAT $(S, \mathcal{I})$ IS A MATROID.

PROOF:

OBVIOUSLY S IS FINITE AND NON-EMPTY, AND $\mathcal{I}$ IS A COLLECTION OF SUBSETS OF S, SO THE FIRST AXIOM IS SATISFIED.

IF $B \subseteq A \in \mathcal{I}$, THEN THE SAME SCHEDULE IN WHICH THE TASKS OF A ARE EARLY ALSO HAS THE TASKS IN B EARLY SINCE $B \subseteq A$. THUS THE HEREDITARY PROPERTY IS SATISFIED.

TO PROVE THE EXCHANGE PROPERTY, LET $A, B \in \mathcal{I}$ WITH $|B| > |A|$. WE MUST SHOW B CONTAINS A TASK WHICH EXTENDS A.

DEFINE

$$k = \max\left\{ t \mid 0 \le t \le n \text{ AND } N_t(B) \le N_t(A) \right\}.$$

RECALL $N_0(B) = N_0(A) = 0$ SO THE ABOVE SET IS NON-EMPTY, WHENCE ITS MAXIMUM k EXISTS. ALSO NOTE

$$N_n(B) = |B| > |A| = N_n(A)$$

SO THAT $k < n$. THE DEFINITION OF k SAYS THAT $N_k(B) \le N_k(A)$ AND

$$N_t(B) > N_t(A) \quad \text{FOR} \quad k < t \le n.$$

IN PARTICULAR

$$N_{k+1}(B) > N_{k+1}(A).$$

THUS

$$N_{k+1}(B) - N_k(B) > N_{k+1}(A) - N_k(A).$$

THIS LAST INEQUALITY SAYS THAT B CONTAINS MORE TASKS WITH DEADLINE $K+1$ THAN DOES A.

LET $a_i \in B - A$ WITH $d_i = K+1$, AND DEFINE

$$A' = A \cup \{a_i\}.$$

WE USE PART (2) OF THE PRECEDING LEMMA TO SHOW $A' \in \mathrm{I}$.

BY THE DEFINITION OF A' WE HAVE $N_t(A') = N_t(A)$ FOR $0 \le t \le K$, AND SINCE $A \in \mathrm{I}$ WE HAVE $N_t(A) \le t$. THUS

$$N_t(A') \le t \quad \text{FOR} \quad 0 \le t \le K.$$

AGAIN BY THE DEFINITION OF A' WE HAVE $N_t(A') \le N_t(A) + 1$ FOR ANY t. BUT RECALL $K < t \le n$ IMPLIES $N_t(A) < N_t(B)$ WHENCE $N_t(A) + 1 \le N_t(B)$. ALSO $N_t(B) \le t$ SINCE $B \in \mathrm{I}$. THUS

$$N_t(A') \le t \quad \text{FOR} \quad K < t \le n.$$

$\therefore A' \in \mathrm{I}$ AS REQUIRED. ///