5. Greedy - Algorithms

= Class of algorithms for optimization problems that always choses in each step a "local" solution that looks best at this moment.

They don't provide always an optimal solution, - but in some case they do (!) - and can also be used as heuristics.

We have also used a greedy-algorithm:

```
DIJKSTRA (G, w, s)
    Init_single_source (G, s)
    S = ∅
    Q := V(G)
    WHILE (Q ≠ ∅)
        u = Extract_MIN (Q)
        S = S ∪ {u}
        FOR (all v ∈ N⁺(u))
            RELAX (u, v, w)
```

"greedy choice": choose element whose "estimated" distance to s is (so-far) the smallest one.

Further Example:

VCP: Find for a graph $G = (V, E)$ a vertex cover of smallest size, i.e., $C \subseteq V$ s.t. $C \cap e \neq \emptyset \ \forall e \in E$.

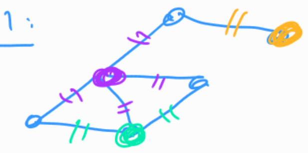
The decision version is NP-complete!

How could a simple heuristic as greedy-alg. look like?

GREEDY-VC (G = (V, E))

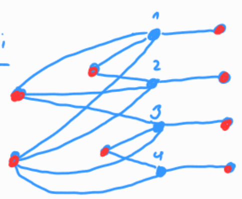
```
Put C = ∅ , E' = E
WHILE ( E' ≠ ∅ )
    v = vertex of max-degree in G' = (V, E')
    C = C ∪ {v}
    remove all edges incident to v from E'
return C
```

Exmpl:



There are instance of VCP for which GREEDY-VC performs very "bad", ie, we are "far" away from an optimal solution.

Exmpl:



$\Rightarrow$ let's have a look to problems for which greedy-algorithms work optimal (can we characterize such type of problems?)

The minimum spanning tree problem

Recap: tree = connected, acyclic graph

forest = graph whose connected components are trees

spanning tree of G is subgraph $T \subseteq G$

s.t. $V(T) = V(G)$

& T is tree

T is tree $\Leftrightarrow$ T connected &

$E(T) = V(T) - 1$.

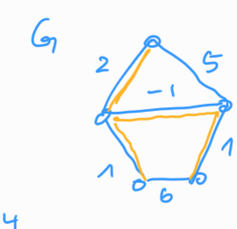
Minimum-Spanning-Tree Problem (MST)

Input: connected graph G with weighting $w: E \to \mathbb{R}$

Aim: Find spanning tree $T \subseteq G$ s.t.

$$w(T) := \sum_{e \in E(T)} w(e) \xrightarrow{!} \min$$

Exmpl:



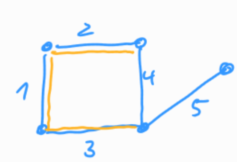
MST T with $w(T) = 3$

Kruskal's Algorithm

KRUSKAL ($G = (V, E)$, $w: E \to \mathbb{R}$) // $m = |E|$

```
SORT edges st: w(e_1) ≤ w(e_2) ≤ ... w(e_m)
F = ∅
T = (V, F)
FOR (i = 1 ... m) DO
    IF [ (V, F ∪ {e_i}) is acyclic ]
        T = (V, F ∪ {e_i})
return T
```

Exmpl



order of edges clear.

at step $i = 4$

we have  + e_4 has cycle

$\to$ goto $i = 5$ ...

- **RUNTIME** [sketch]:

Sort edges : $O(|E| \log(|E|)) = O(|E| \log |V|)$ since $|E| \leq |V|^2$ $\Rightarrow \log(|E|) \leq \log(|V|^2) = 2 \log(|V|)$

$F = \emptyset$: $O(1)$

$T = (V, F)$: $O(|V|)$

FOR loop : in the i-th step, you only need to consider those subgraphs of G, whose vertices are contained in the edges in T chosen so-far

$\Rightarrow$ save as auxil. graph A — has at most $2i$ vertices and i edges in step i

in the aux-graph A test "acyclic" via Breadth first search (BFS) in $O(|E(A)| + |V(A)|)$

$= O(i + 2i) = O(i) \leq O(n)$, $n = |V|$, $m = |E|$

m steps, each $O(n)$ $\Rightarrow$ $O(m \cdot n)$

Total: $O(|E||V|)$

with efficient datastructure, this can be improved to $O(|E| \log(|V|))$

- **Correctness?**

Theorem 5.1: Kruskal's algorithm correctly computes a MST for given undirected graph G.

proof: 1) Alg. terminates ✓

2) resulting graph T contains is acyclic & $V(T) = V(G)$ } T is spanning forest of G.

3) T is connected: IF not $\Rightarrow \exists x, y \in V(T)$ st no x-y-path exists in T

G is connected $\Rightarrow \exists$ x-y-path in G.

$\Rightarrow$ x-y-path & edge on e at $e \notin F$: 

at some i-th step of FOR-loop edge e is considered.

But $T' = (V, F \cup \{e\})$ remains acyclic (otherwise there would have been a x-y-path already in T' & thus in T)
(at step i)

$\Rightarrow$ e is added to T at step i $\lightning$ $e \notin F$

$\Rightarrow$ T is spanning tree of G.

remains to show that T is <u>minimum</u> spanning tree of G.

Let F_i be the set of edges formed up to step i

<u>Claim:</u> $\exists$ MST T^* of G st $F_k \subseteq E(T^*)$.

$i = 0$, $T = (V, \emptyset)$ ✓

<u>Ind hyp:</u> Assume claim is true for all Steps $i \leq k$

Let T^* be MST of G st $F_k \subseteq E(T^*)$

now: $i = k+1$. and let $e = (uv)$ be the current edge considered in Step $i = k+1$, i.e.,

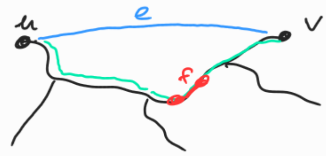
- IF e not added to F_k
  THEN $F_{k+1} = F_k \subseteq E(T^*)$ ✓
- IF e added to F_k
  THEN $F_{k+1} = F_k \cup \{e\}$

$\hookrightarrow$ IF $e \in E(T^*) \Rightarrow F_{k+1} \subseteq E(T^*)$ ✓

ELSE $e \notin E(T^*)$ & therefore, $(V, E^*(T) \cup \{e\})$ must contain a cycle.

[↗ Thm 1.1.(4)]

Situation: T^*



u-v-path P in T^*

$e = (uv)$

Since e is contained in T, at least one of edge f to P cannot be contained in T, otherwise T contains cycle.

$\Rightarrow f \notin F_k \subseteq E(T)$

Now, $\tilde{T} =$ "$T^* - f + e$" is again a tree.

Moreover, $w(f) \geq w(e)$ [otherwise if $w(f) < w(e)$, then the greedy-choice would be f & not e]

$\Rightarrow w(\tilde{T}) \leq w(T^*)$

& since T^* is MST $\Rightarrow w(\tilde{T}) \geq w(T^*)$ $\Rightarrow$ $w(\tilde{T}) = w(T^*)$

$\Rightarrow$ $\tilde{T}$ is MST that contains the edges in F_{k+1}.

Now repeat the latter arguments until all edges of are contained in MST

$\square$

The coin-exchange problem

Situation: you are paying cash & only coins will be returned.

Aim: you want to get back the least number of coins!

SEK: 10 kr, 5 kr, 2 kr, 1 kr.

Solution: Greedy $\rightarrow$ return largest valued-coins as long as possible. "greedy-choice"

Works also for €, $, but not for all!

Exmpl: Coins: 1, 3, 4 & return value 6

$\Rightarrow$ greedy 4, 1, 1
instead of 3, 3.

Lemma 5.2 One minimizes the number of returned coins for SEK $c_1 = 1, c_2 = 2, c_3 = 5, c_4 = 10$, when using the "greedy choice" for some return value X

proof: enumeration

X	1	2	3	4	5	6	7	8	9
C	1	2	2,1	2,2	5	5,1	5,2	5,2,1	5,2,2

optimal.

let $X \geq 10$

Assume $c_4 = 10$ & C^* optimal solution

$\Rightarrow X$ can be returned by 5, 2, 1 values

$\Rightarrow X = k \cdot 5 + r$, k max $\Rightarrow r$ must be returned in 1, 2 coins. alues

as above $\forall y$ with $5 \leq y < 10$

there is a coin $c_3 = 5$

$\Rightarrow r = x - 5k < 5$ (otherwise k not max)

$\Rightarrow \quad 10 \leq x < 5(k+1) \Rightarrow k \geq 2$

$\Rightarrow$ there are at least 2 coins $c_3 = 5$ that one can replace by $c_4 = 10$ to obtain "better" solution than C^* $\lightning$

$\square$

We have seen now a couple of examples, in some cases greedy works well, in some cases not $\Rightarrow$ can we characterize the class of problems, where greedy returns always an optimal solution?

Answer: YES; matroids.

5.1. Matroids

Optimization problems can usually be formulated as:
- groundset E,
- set $\mathbb{F}$ of (subsets of) feasible solutions st $\mathbb{F} \subseteq \mathbb{P}(E)$
  powerset of E
- weighting $w: \mathbb{F} \to \mathbb{R}$

Aim: find inclusion-maximal element $F \in \mathbb{F}$ st $w(f) \to \min/\max$

Exmpl: MST here $E = E(G)$ of given graph $G = (V, E)$
$\mathbb{F} = \{ E' \subseteq E : (V, E') \text{ acyclic } (= \text{Forest}) \}$

span that contains all spanning trees & its subgraphs (from which we construct the final solution)

$F \in \mathbb{F}$ inclusion-maximal $\Rightarrow$ (V, F) spanning tree $+ \; w(F) \to \min \Rightarrow (V, F)$ MST.

Def: Let E be a finite set, $\mathbb{F} \subseteq \mathbb{P}(E)$

Then, $(E, \mathbb{F})$ is called independent system

if (M1) $\emptyset \in \mathbb{F}$

& (M2) $X \subseteq Y,\ Y \in \mathbb{F} \to X \in \mathbb{F}$ [closed wrt. $\subseteq$]

An independent system $(E, \mathbb{F})$ is a matroid if it satisfies

(M3) $X, Y \in \mathbb{F}$ & $|X| > |Y|$ [Exchange property]

$\Rightarrow \exists\, x \in X \setminus Y$ st $Y \cup \{x\} \in \mathbb{F}$

Elements in $\mathbb{F}$ are called <u>independent</u>
Elements in $\mathcal{P}(E) \setminus \mathbb{F}$ are <u>dependent</u>

Let $\mathbb{M} \subseteq \mathcal{P}(E)$,
Element $F \in \mathbb{M}$ is (inclusion-) <u>maximal</u>
if $\nexists F' \in \mathbb{M}$ st $F \subsetneq F'$
Element $F \in \mathbb{M}$ is (inclusion-) <u>minimal</u>
if $\nexists F' \in \mathbb{M}$ st $F' \subsetneq F$

minimal dependent sets [elements from $\mathcal{P}(E) \setminus \mathbb{F}$]
are called <u>cycles</u> of $\mathbb{F}$
[NOTE all subsets of cycles are dependent]

maximal independent sets [elements from $\mathbb{F}$]
are called <u>basis</u> of $\mathbb{F}$

The notion of "independence" above generalizes the term „linear independence" in Vectorspaces to arbitrary combinatorial objects.

<u>Exmpl</u>: vector $e_1 = \begin{pmatrix}1\\0\end{pmatrix}$, $e_2 = \begin{pmatrix}0\\1\end{pmatrix}$, $e_3 = \begin{pmatrix}0\\0\end{pmatrix}$,
$E = \{e_1, e_2, e_3\}$
$\mathbb{F} = \{ F \subseteq E : \text{vectors in } F \text{ are lin. indep.}\}$
$\Rightarrow \mathbb{F} = \{\emptyset, \{e_1\}, \{e_2\}, \{e_1, e_2\}\}$ // 0-vector always lin. depend.

$\mathcal{P}(E) \setminus \mathbb{F}$ = sets of vectors that are lin. depend.
$= \{e_3, \{e_1, e_3\}, \{e_2, e_3\}, \{e_1, e_2, e_3\}\}$
Cycles: e_3 , Basis: $\{e_1, e_2\}$

Further Examples of matroids

(a) <u>trivial matroids</u> $(E, \mathbb{F} = \{\emptyset\})$

Basis $\emptyset$, cycles are in $\mathbb{P}(E) \setminus \{\emptyset\}$ all (inclusion)-min. subsets of E

<u>proof</u>: (M1): $\emptyset \in \mathbb{F}$ ✓

(M2): $X \subseteq Y \in \mathbb{F}$, since $\mathbb{F} = \{\emptyset\} \Rightarrow X = Y = \emptyset \in \mathbb{F}$ ✓

(M3): trivial since $|\mathbb{F}| = 1$ ✓

(b) <u>free matroids</u> $(E, \mathbb{P}(E))$

Basis all incl. max elements of E, no cycles since $\mathbb{P}(E) \setminus \mathbb{P}(E) = \emptyset$.

(M1): $\emptyset \in \mathbb{F} = \mathbb{P}(E)$ ✓.

(M2): $X \subseteq Y \in \mathbb{F} \Rightarrow X \in \mathbb{P}(E) = \mathbb{F}$ ✓

(M3): clear ✓

(c) <u>uniform matroid</u> $(E, \mathbb{F})$ with

E finite set, $|E| \geq k \geq 0$

$\mathbb{F} = \{ F \subseteq E : |F| \leq k \}$

Basis: set $\{ F \subseteq E : |F| = k \}$, cycles in $\mathbb{P}(E) \setminus \mathbb{F}$ = set of elements $F \subseteq E$ with $|F| > k$

cycle those with $|F| = k+1$

(M1) $|\emptyset| = 0 \leq k \Rightarrow \emptyset \in \mathbb{F}$ ✓

(M2) $X \subseteq Y \in \mathbb{F} \Rightarrow |X| \leq |Y| \leq k \Rightarrow X \in \mathbb{F}$

(M3) $X, Y \in \mathbb{F}$, $k \geq |X| > |Y| \Rightarrow k-1 \geq |Y|$

$\Rightarrow$ take $x \in X \setminus Y \to |\underbrace{Y \cup \{x\}}_{\in \mathbb{F}}| \leq k$ ✓

(d) <u>Vector matroid</u>: E = set of vectors

$\mathbb{F} = \{ F \subseteq E :$ vectors in F lin. indep. $\}$.

(M1): $\emptyset \in \mathbb{F}$ ✓

(M2): $X \subseteq Y \in \mathbb{F} \Rightarrow$ vect. in X lin. ind. $\Rightarrow X \in \mathbb{F}$

(M3): Steinitz exchange lemma (Lin. Alg)

Basis: "algebraic" basis

(e) <u>cycle matroid</u> given graph $G = (V, E)$

(if loops are allowed: graphic matroid)

$\mathbb{F} = \{ F \subseteq E : (V, F) \text{ is a forest} \}$.

Basis: F with (V, F) tree.

<u>Lemma 5.3:</u> $(E, \mathbb{F})$ as defined in (e) is a matroid.

<u>proof:</u> (M1) $(V, \emptyset)$ Forest $\Rightarrow \emptyset \in \mathbb{F}$ ✓

(M2) $X \subseteq Y \in \mathbb{F} \Rightarrow (V, Y)$ is Forrest,
& $(V, X) \mathrel{\hat{=}} (V, Y)$ - some edges
remains forest ✓

Assume, for contradiction, that (M3) is not satisfied.

$\Rightarrow \exists\ X, Y \in \mathbb{F}$ st $|X| > |Y|$ but for all $x \in X \setminus Y$
we have $(V, Y \cup \{x\})$
is not a forest

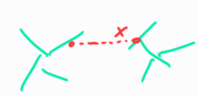
$Y \in \mathbb{F} \Rightarrow (V, Y)$ is forest, let $x \in X \setminus Y$

$\Rightarrow (V, Y \cup \{x\})$ contains cycles
in one of the trees + edge x in (V, Y)

conn. comp. of (V, Y):

Let T be a connected component of (V, X)
(= tree)

$\Rightarrow$ each edge of T is either contained in Y
[and thus, part of a CC of (V, Y)]
or "closes" a cycle.

($\Rightarrow$ this is not possible: )

$\Rightarrow$ we have either: 

$\Rightarrow T$ is entirely contained in one
conn. comp. T' of (V, Y)

$\Rightarrow V(T) \subseteq V(T')$

$\Rightarrow$ every tree of (V, X) is entirely contained in some tree
of (V, Y)

$\Rightarrow$ $p := \#$ trees of $(V, X) \geq \#$ trees of $(V, Y) =: q$

we know: # edges in tree T: $|E(T)| = |V(T)| - 1$

$\overset{\text{Exerc}}{\Longrightarrow}$ # edges in forest F: $|E(F)| = |V(F)| - k$
with k conn. comp

$\Rightarrow |V| - |X| = p \geq q = |V| - |Y| \Rightarrow |Y| \geq |X|$ ↯ $|X| > |Y|$

$\Rightarrow$ (M3) must be satisfied $\square$

Reasonable question: Why are we doing this?
Answer: to characterize problems
on which greedy always
returns optimal solution!

To this end, we need some further results.

<u>Theorem 5.4</u> All basiselements of a matroid have the same size.

<u>Proof:</u> Let B_1, B_2 be basiselements of $(E, \mathbb{F})$,
i.e. they are <u>maximal</u> elements in $\mathbb{F}$.

Assume, for contradiction, $|B_1| \neq |B_2|$.

W.l.o.g. $|B_1| < |B_2|$

$\overset{(M3)}{\Longrightarrow} \exists\, x \in B_2 \setminus B_1$ st $B_1 \cup \{x\} \in \mathbb{F}$

$\Rightarrow B_1$ was not maximal ↯ $\square$

Proposition 5.5 Let $(E, \mathbb{F})$ be an independent system (satisfies M1 & M2)

The following statements are equivalent:

(M3) $X, Y \in \mathbb{F}$, $|X| > |Y| \Rightarrow \exists\, x \in X \setminus Y$ st $Y \cup \{x\} \in \mathbb{F}$

(M3') $X, Y \in \mathbb{F}$, $|X| = |Y| + 1 \Rightarrow \exists\, x \in X \setminus Y$ st $Y \cup \{x\} \in \mathbb{F}$

proof: (M3) ⇒ (M3'): clear ✓

(M3') ⇒ (M3): Let $X, Y \in \mathbb{F}$, $|X| > |Y|$

$\overset{(M2)}{\Rightarrow} \forall X' \subseteq X : X' \in \mathbb{F}$

$\Rightarrow \exists\, X' \subseteq X$ st $X' \in \mathbb{F}$ & $|X'| = |Y| + 1$

$\overset{(M3')}{\Rightarrow} \exists\, x \in X' \setminus Y \subseteq X \setminus Y$ st $Y \cup \{x\} \in \mathbb{F}$

□

Max/min problems:

Given: independ. system $(E, \mathbb{F})$ + weight $w : E \to \mathbb{R}_{\geq 0}$

Aim: Find element $X \in \mathbb{F}$ st $w(X) = \sum_{e \in X} w(e) \overset{!}{\longrightarrow}$ min/max

GREEDY$(E, \mathbb{F})$ [for max; min analog but increasing order]

```
SORT E st w(e1) ≥ w(e2) ≥ ... ≥ w(em)
F = ∅
FOR (i=1 ... m) DO
    IF (F ∪ {ei} ∈ F)
        F = F ∪ {ei}
RETURN F
```

// eg. Kruskal $F \cup \{e_i\} \in \mathbb{F}$ if $(V, F \cup \{e_i\})$ acyclic

Runtime: sort: $O(m \log m)$ depends on specific problem, e.g. "test if acyclic"

FOR: m times

test $F \cup \{e_i\} \in \mathbb{F}$ in $f(n)$ time

$\Rightarrow \quad O(m \log m + m \cdot f(n))$ time.

Theorem 5.6 (Edmonds-Rado Thm): Let $(E, \mathbb{F})$ be an indep. syst.

Then: $(E, \mathbb{F})$ is matroid $\Leftrightarrow$ GREEDY finds optimal solution f. a possible weights $w: E \to \mathbb{R}_{\geq 0}$

proof:

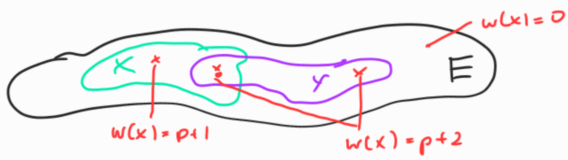
"$\Leftarrow$" By contraposition, assume $(E, \mathbb{F})$ is not a matroid

$\overset{\text{Prop 5.5}}{\Longrightarrow}$ $(M3')$ is not satisfied.

$\Rightarrow \exists \; X, Y \in \mathbb{F}, \quad |X| = p+1, \; |Y| = p$

st $\forall \; x \in X \setminus Y: \quad Y \cup \{x\} \notin \mathbb{F}$

Let $w: E \to \mathbb{R}_{\geq}$ st $w(x) = \begin{cases} p+2, & x \in Y \\ p+1, & x \in X \setminus Y \\ 0, & \text{else} \end{cases}$



$$\Rightarrow w(X) = \sum_{x \in X} w(x) = \sum_{x \in X \setminus Y} p+1 + \sum_{x \in X \cap Y} p+2$$

$$\geq |X| \cdot (p+1) = (p+1)^2 = p^2 + 2p + 1 > p^2 + 2p = p \cdot (p+2) = w(Y)$$

GREEDY will first chose all elements from Y

since $w(y) > w(x) \; \forall y \in Y, x \in X \setminus Y$

$\Rightarrow$ GREEDY-solution: Y + "smth", but "smth" cannot be in X & thus, has weight 0

$\Rightarrow w(Y)$ is returned opt. solution

$\Rightarrow$ GREEDY does not find optimal solution.

„$\Rightarrow$"

Let $(E, \mathcal{F})$ be a matroid.

Let GREEDY solution be: $I = \{e_{l_1} \dots e_{l_i}\} \in \mathcal{F}$ $\quad w(e_1) \geq \dots \geq w(e_i)$

Let $J = \{f_{r_1} \dots f_{r_j}\}$ be other feasible solution, $\quad w(f_1) \geq \dots \geq w(f_j)$
s.t. J is basis (= max. indep. element) of $\mathcal{F}$.

Note I is a basis (= max indep) of $\mathcal{F}$, since otherwise GREEDY would have added more elements (since $w(x) \geq 0 \ \forall x \in E$)

$\Rightarrow |I| = |J|$ i.e. $i = j$.

We show: $\forall m \in \{1, 2, \dots i\} : w(e_{l_m}) \geq w(f_{r_m})$

Assume, for contradiction, this is no true

$\Rightarrow \exists$ smallest index k s.t. $w(e_{l_k}) < w(f_{r_k})$

Consider $I' = \{e_{l_1} \dots e_{l_{k-1}}\}$
$J' = \{f_{r_1} \dots f_{r_{k-1}}, f_{r_k}\}$ & thus, $|I'| < |J'|$

$\overset{(M3)}{\Rightarrow} \exists f_s \in J' \setminus I'$ s.t. $I' \cup \{f_s\} \in \mathcal{F}$

Note $w(f_s) \geq w(f_{r_k}) > w(e_{l_k})$

by ordering $w(f_1) \geq \dots \geq w(f_k)$ — some of this is $w(f_s)$ $\quad$ by assumpt.

$w(f_{r_t}) \geq w(f_{r_k}) > w(e_{l_k})$

$\forall f_{r_t} \in J' \setminus I'$

$\Rightarrow$ that is f_s is "ordered before" e_k & GREEDY would have chosen f_s instead of e_k, but then $f_s \in I'$ ↯ $f_s \in J' \setminus I'$

$\Rightarrow \forall m \in \{1, 2, \dots i\} : w(e_{l_m}) \geq w(f_{r_m})$

$\Rightarrow w(I) \geq w(J) \quad \forall$ possible feasible solution

$\Rightarrow I$ is an optimal solution $\quad \square$

Summary

Matroids characterize a class of optimization problems that can be solved by greedy alg. assuming $w(e) \geq 0 \; \forall e \in E$

The latter can be generalized to all "well-behaved" weighting functions via so called greedoids that generalize the notion of matroids.
Greedoids just satisfy M1 & M3. & so M2 generalizes to: $\forall X \in \mathbb{F} \; \exists x \in X$ s.t. $X \setminus \{x\} \in \mathbb{F}$.

$\Rightarrow$ every matroid is a greedoid.

The notion of matroids can be used to show that for certain problems greedy works.

But in a similar way one can show that greedy does not work:

E.g. HAMILTONIAN CYCLE

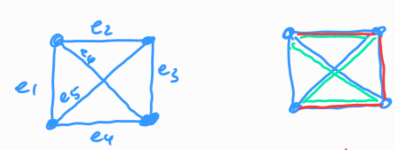
// NP complete $\Rightarrow$ greedy doesn't work, but one can show this directly.

create indep. system for given graph G:

$E = E(G)$

$\mathbb{F} = \{ E' \subseteq E \mid E'$ subset of edges of Ham. cycle in G $\}$

(M1), (M2) ✓
(M3) ?



$Y = \{e_2, e_3, e_4\}$
$X = \{e_2, e_4, e_5, e_6\}$

$\Rightarrow \forall e \in X \setminus Y = \{e_5, e_6\}: Y \cup \{e\} \notin \mathbb{F}$!
$\Rightarrow$ so (M3) is in general not satisfied.