Greedy Set Cover

Set Cover$(U, \mathcal{S})$

$\mathcal{C} \leftarrow \emptyset$
while some elements of U are not covered by $\mathcal{C}$
  do
    Find set S with maximum cost-effectiveness
      where cost-effectiveness of S is cost(S)/(number of not-yet-covered elements of S)
    Add S to C

return $\mathcal{C}$

Cost-effectiveness is a quantity that depends on the set but also on the sets previously chosen. It changes as more and more elements are covered. Note, however, that it can never decrease as we add more and more sets to the cover.

Given a sequence of sets forming a set cover $\mathcal{C}$, for each element e_k we can define a price, set to be the cost-effectiveness of the first set S in the sequence that covers e_k, at the time when S is put in the cover. It is easy to see that the sum of element prices equals the total cost of $\mathcal{C}$.

Theorem 1 *The above is an H_n-approximation algorithm for Set Cover.*

Label the elements of U $e_1, e_2, \ldots, e_n$ in the order in which they are covered by Greedy. Let U_k denote the elements of U that are not yet covered, right before e_k is covered by Greedy. Note that U_k has size at least $n - (k-1)$. Consider the instance $(U_k, \mathcal{S}_k)$ of Set Cover obtained from $(U, \mathcal{S})$ by intersecting each set S_i pf $\mathcal{S}$ with U_k. Let α be the cost-effectiveness of the set chosen by Greedy to cover e_k (at the time when Greedy chooses it), or in other words, the price of e_k for Greedy. By definition of Greedy, for the instance $(U_k, \mathcal{S}_k)$ initially every set has cost-effectiveness greater than or equal to α.

Let OPT_k be the optimal set cover of U_k. Note that $\text{OPT}_k \leq \text{OPT}$. Take the sets of OPT_k in some arbitrary order, thus defining a sequence, hence a cost-effectiveness for each set of OPT_k, and thus a price for each element of U_k. Since cost-effectiveness can never decrease with time, every set in OPT_k has cost-effectiveness at least α, and so every element has price at least α for OPT_k, and so the cost of OPT_k is at least $\alpha|U_k|$, which is at least $\alpha(n-k+1)$.

Rewriting, the price of e_k for Greedy is $\alpha \leq \text{cost}(\text{OPT}_k)/(n-k+1)$. Summing over k, the total price of all items for Greedy is at most $H_n\text{cost}(\text{OPT}_k)$, which is at most $H_n\text{cost}(\text{OPT})$, hence the Theorem.