Topic 3: Set Cover Problem

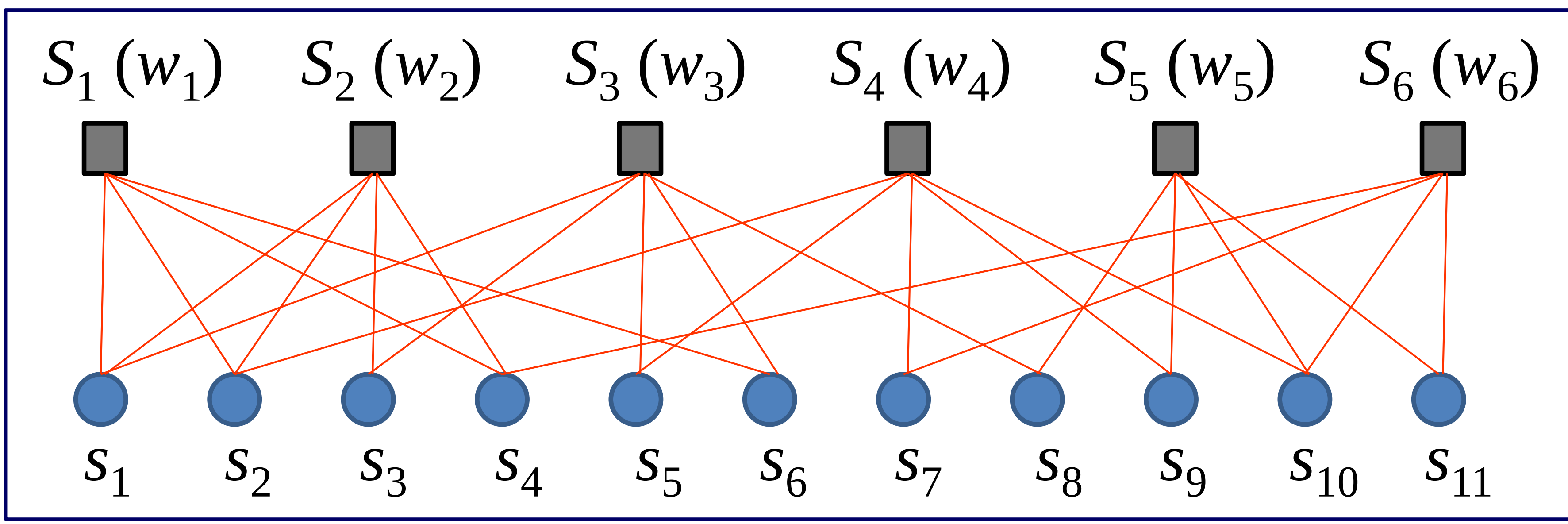
Input: n elements: $U = \{s_1, s_2, ..., s_n\}$

m subsets of U: $S_1, S_2, ..., S_m$ $(S_i \subset U)$

Weight (cost) of each subset: w_i $(i = 1, 2, ..., m)$

Output: Cover C (Selection from m subsets): $\bigcup_{S_i \in C} S_i = U$

Objective: Minimize the total weight: $\sum_{S_i \in C} w_i$



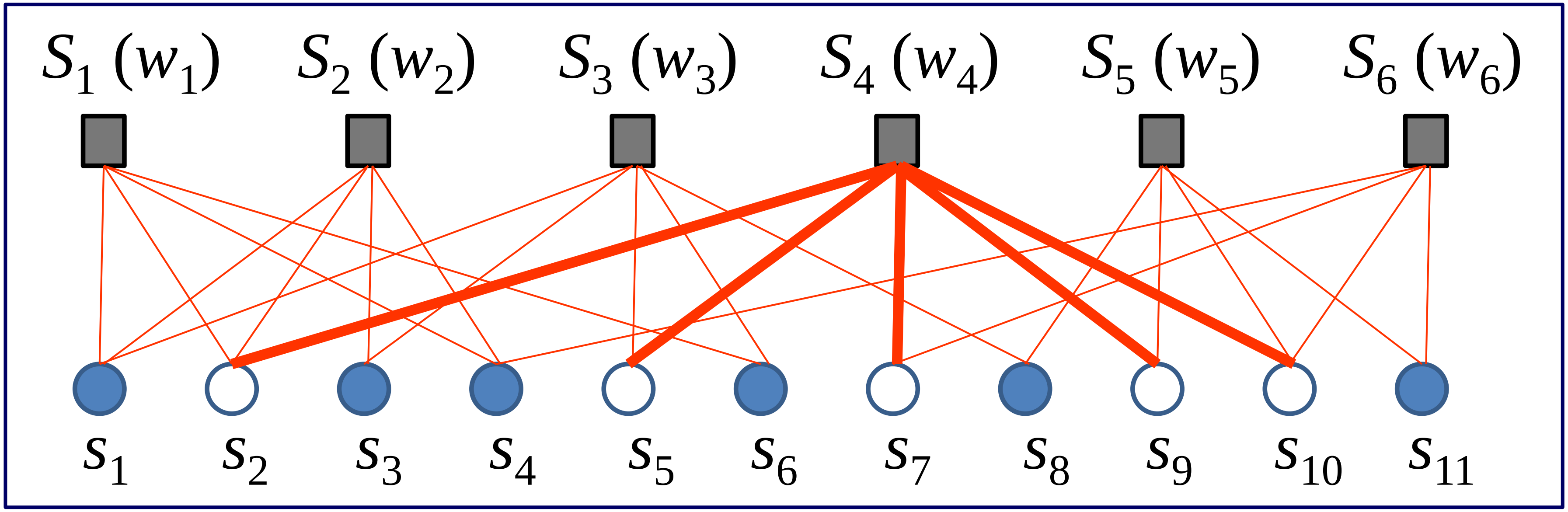
Minimize $w = \sum_{S_i \in C} w_i$ subject to $\bigcup_{S_i \in C} S_i = U$

Good subset: Small weight with many elements $\frac{w_i}{|S_i|}$

After some elements are covered
(*R*: remaining uncovered elements): $\frac{w_i}{|S_i \cap R|}$

Greedy Set Cover Algorithm

Select the best subset with the best evaluation.



Greedy Set Cover Algorithm

Select the best subset with the best evaluation.

(Programming Exercise 6)

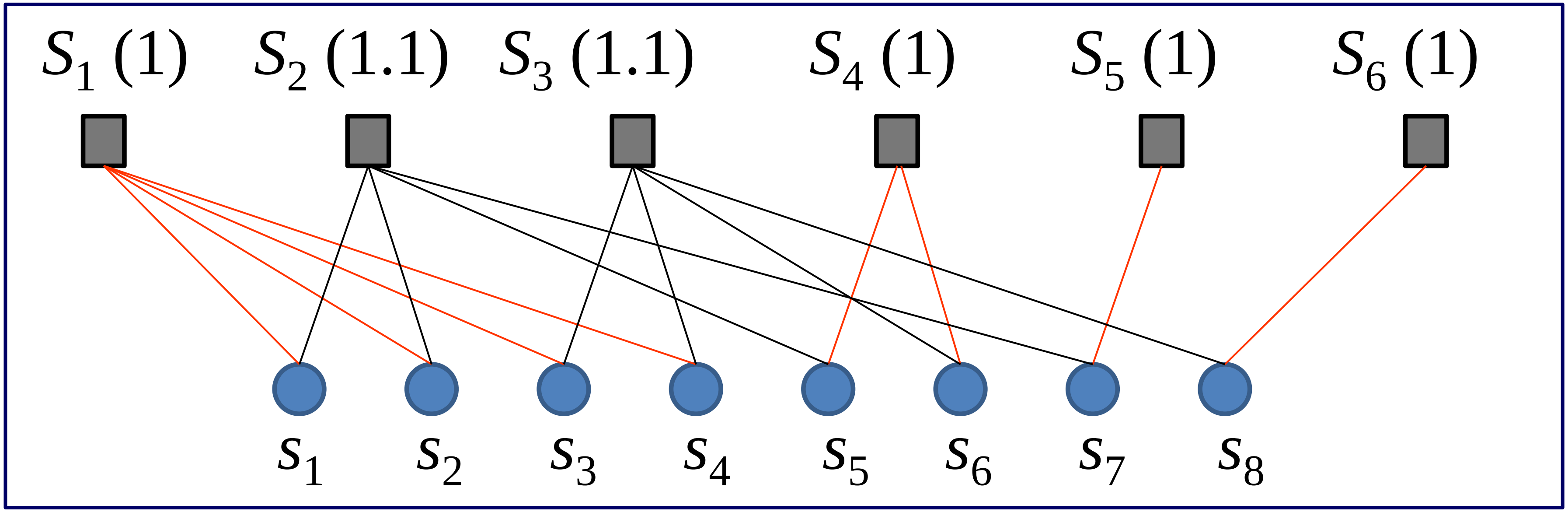
procedure GREEDY-SET-COVER
  Start with $R = U$ and no sets selected
  while $R \neq \emptyset$ **do**
    Select set S_i that minimizes $\frac{w_i}{|S_i \cap R|}$
    Delete set S_i from R
  end while
  Return the selected sets
end procedure

Exercise 6-1:

Create an example of the set cover problem where a good solution is not obtained by the greedy algorithm.

Simple Example

w = 4 by C = {S_1, S_4, S_5, S_6, }, w = 2.2 by C = {S_2, S_3}



Approximation Quality of Algorithm: ?-approximation

When an element s is covered by S_i, the cost c_s paid by s is

$$c_s = \frac{w_i}{|S_i \cap R|} \quad \text{for all } s \in S_i \cap R$$

(since the total cost paid by all elements covered by S_i is w_i.)

If C is the cover obtained by the greedy set cover algorithm and c_s is calculated during the execution of the algorithm,

$$\sum_{S_i \in C} w_i = \sum_{s \in U} c_s$$ (the right-hand side will be evaluated)

Preparation

Harmonic Function: $H(n) = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$

(i) For every set S_k,

(c_s is calculated during the execution of the greedy algorithm)

Let us assume that d elements in $S_k = \{s_1, s_2, ..., s_d\}$ is covered in the order of $s_1, s_2, ..., s_d$ by the greedy algorithm. Consider the iteration when s_j is covered. Before this iteration, $\{s_j, s_{j+1}, ..., s_d\} \subset R$. Thus $\dfrac{w_k}{|S_k \cap R|} = \dfrac{w_k}{d - j + 1}$

At this iteration, the algorithm selects S_i with the minimum average cost. So, $c_{s_j} = \dfrac{w_i}{|S_i \cap R|} \leq \dfrac{w_k}{|S_k \cap R|} = \dfrac{w_k}{d - j + 1}$

Thus

$$\sum_{s \in S_k} c_s = \sum_{j=1}^{d} c_{s_j} \leq \sum_{j=1}^{d} \frac{w_k}{d - j + 1} = \frac{w_k}{d} + \frac{w_k}{d-1} + \cdots + \frac{w_k}{1} = w_k H(d)$$

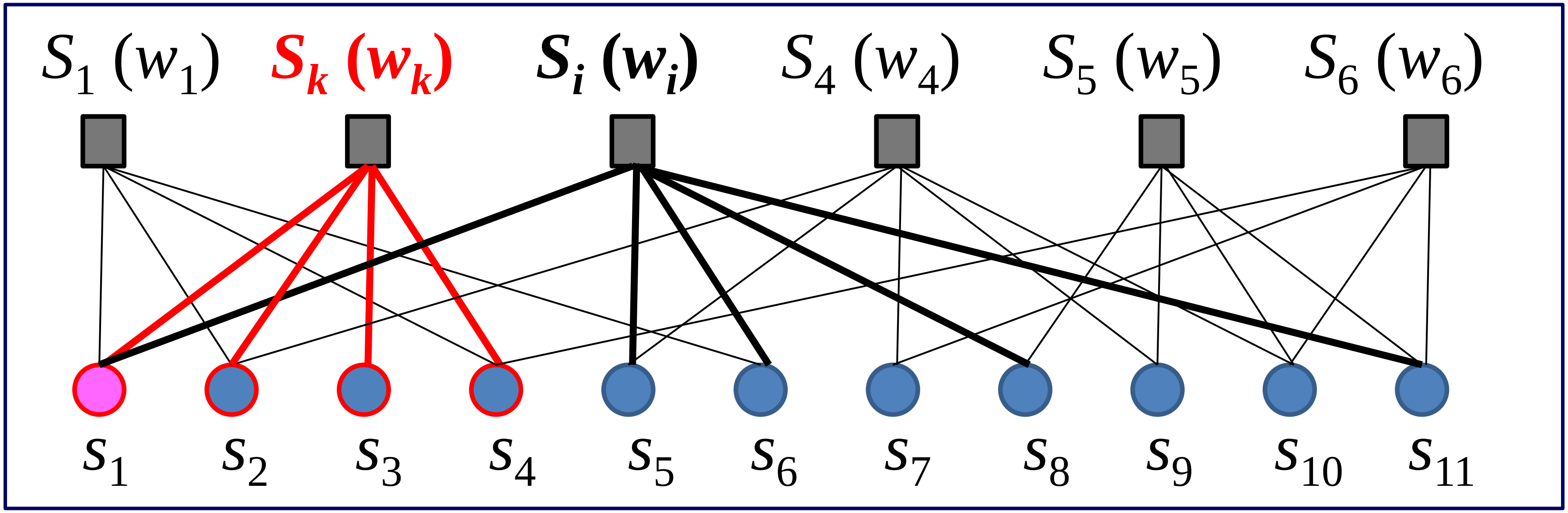
Example

If s_1 is covered by S_i (not S_k), the following relation holds:

$$c_1 = \frac{w_i}{|S_i \cap R|} \le \frac{w_k}{|S_k \cap R|} = \frac{w_k}{d-j+1} = \frac{w_k}{4} \quad (d=4,\ j=1)$$

If s_1 is covered by S_k, the following relation holds:

$$c_1 = \frac{w_k}{|S_k \cap R|} = \frac{w_k}{d-j+1} = \frac{w_k}{4}$$



$$S_k = \{s_1, s_2, s_3, s_4\}$$

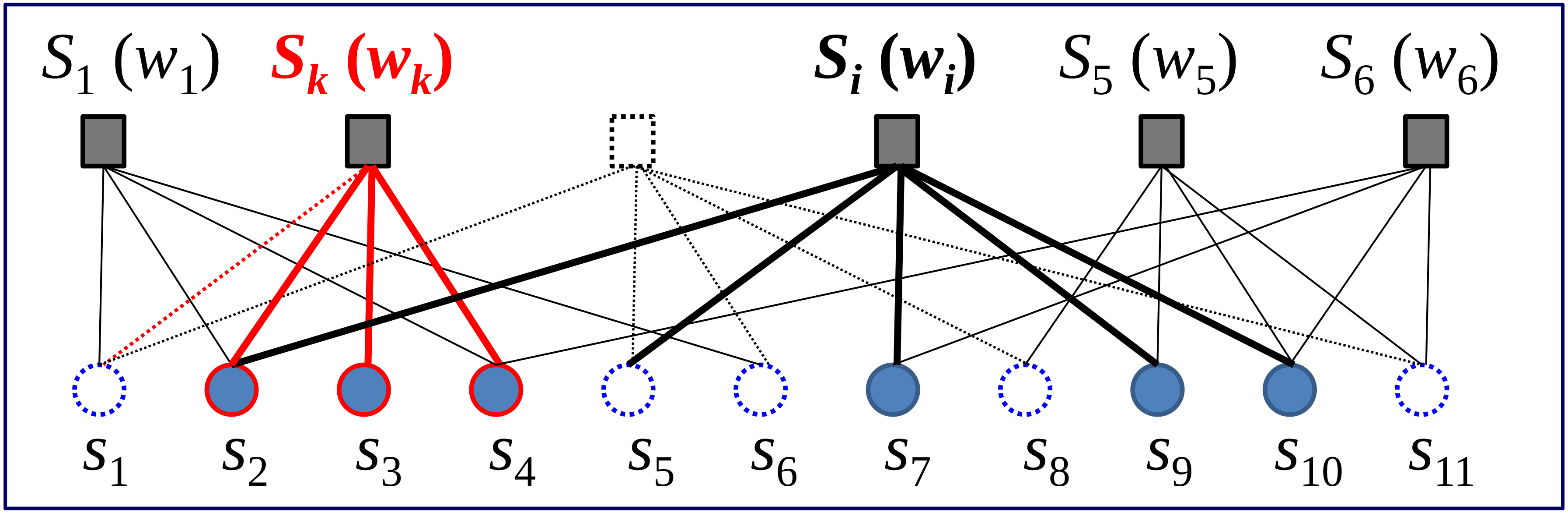
Example

If s_2 is covered by S_i (not S_k), the following relation holds:

$$c_2 = \frac{w_i}{|S_i \cap R|} \le \frac{w_k}{|S_k \cap R|} = \frac{w_k}{d-j+1} = \frac{w_k}{3} \quad (d = 4,\ j = 2)$$

If s_2 is covered by S_k, the following relation holds:

$$c_2 = \frac{w_k}{|S_k \cap R|} = \frac{w_k}{d-j+1} = \frac{w_k}{3}$$



$$\sum_{s \in S_k} c_s = \sum_{j=1}^{d} c_{s_j} \le \sum_{j=1}^{d} \frac{w_k}{d-j+1} = \frac{w_k}{d} + \frac{w_k}{d-1} + \cdots + \frac{w_k}{1} = w_k H(d)$$

(ii) $w \le H(\max_k |S_k|)\, w^*$

The obtained weight ***w*** by the greedy algorithm is not worse than $\boldsymbol{H(d^*)}$ times of the optimal weight $\boldsymbol{w^*}$ where $\boldsymbol{d^*} = \max_{\boldsymbol{k}} |\boldsymbol{S_k}|$.

Let C^* be the optimal set cover: $w^* = \sum_{S_i \in C^*} w_i$

From (i), we have

$$\sum_{s \in S_i} c_s \le H(|S_i|) w_i \le H(d^*) w_i \;\Rightarrow\; w_i \ge \frac{1}{H(d^*)} \sum_{s \in S_i} c_s$$

Since C^* is a set cover, $\sum_{S_i \in C^*} \sum_{s \in S_i} c_s \ge \sum_{s \in U} c_s$

Thus

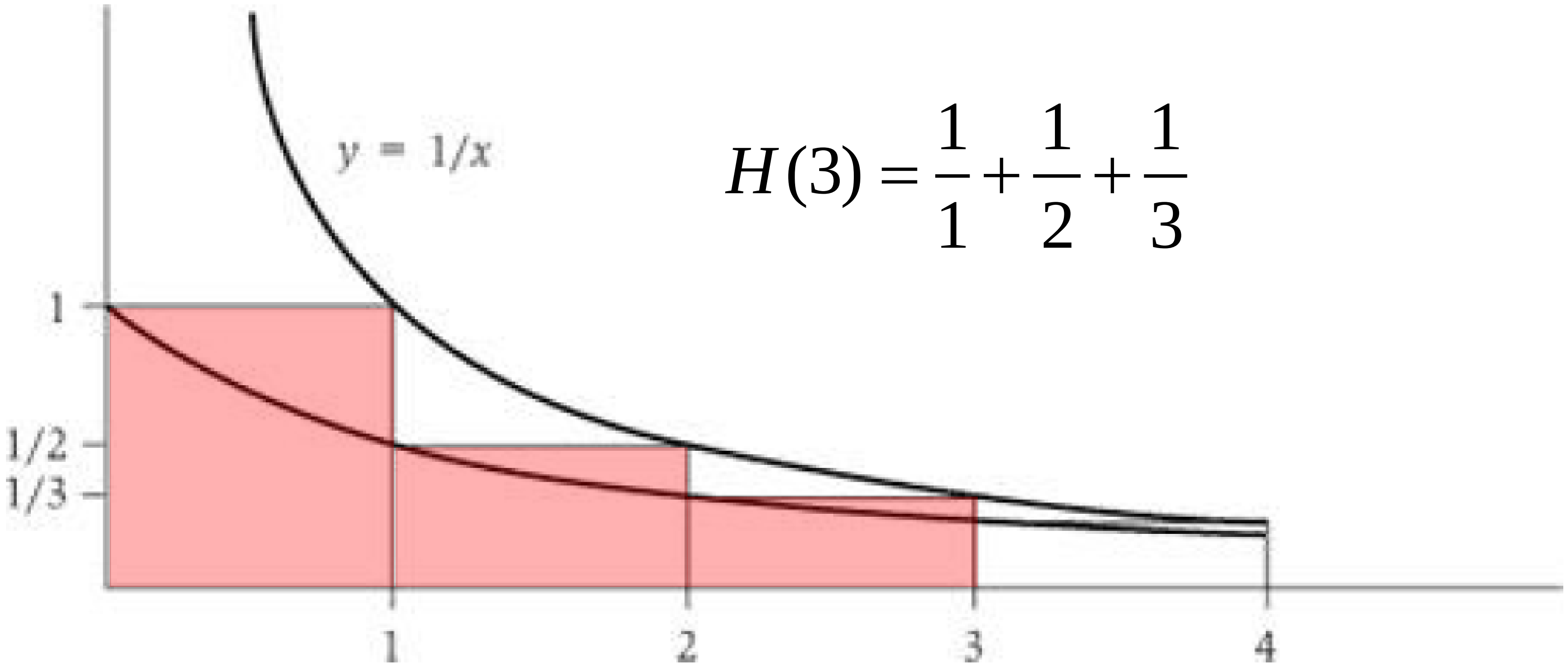
$$w^* = \sum_{S_i \in C^*} w_i \ge \sum_{S_i \in C^*} \left[\frac{1}{H(d^*)} \sum_{s \in S_i} c_s \right] \ge \frac{1}{H(d^*)} \sum_{s \in U} c_s = \frac{1}{H(d^*)} \sum_{S_i \in C} w_i$$

$$w^* \ge \frac{w}{H(d^*)} \;\Rightarrow\; w \le H(d^*)\, w^*$$

Harmonic Function: $H(n) = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$

$$\int_1^{n+1} (1/x)dx = \ln(n+1) \;\leq\; H(n) \;\leq\; 1 + \int_1^n (1/x)dx = 1 + \ln(n)$$

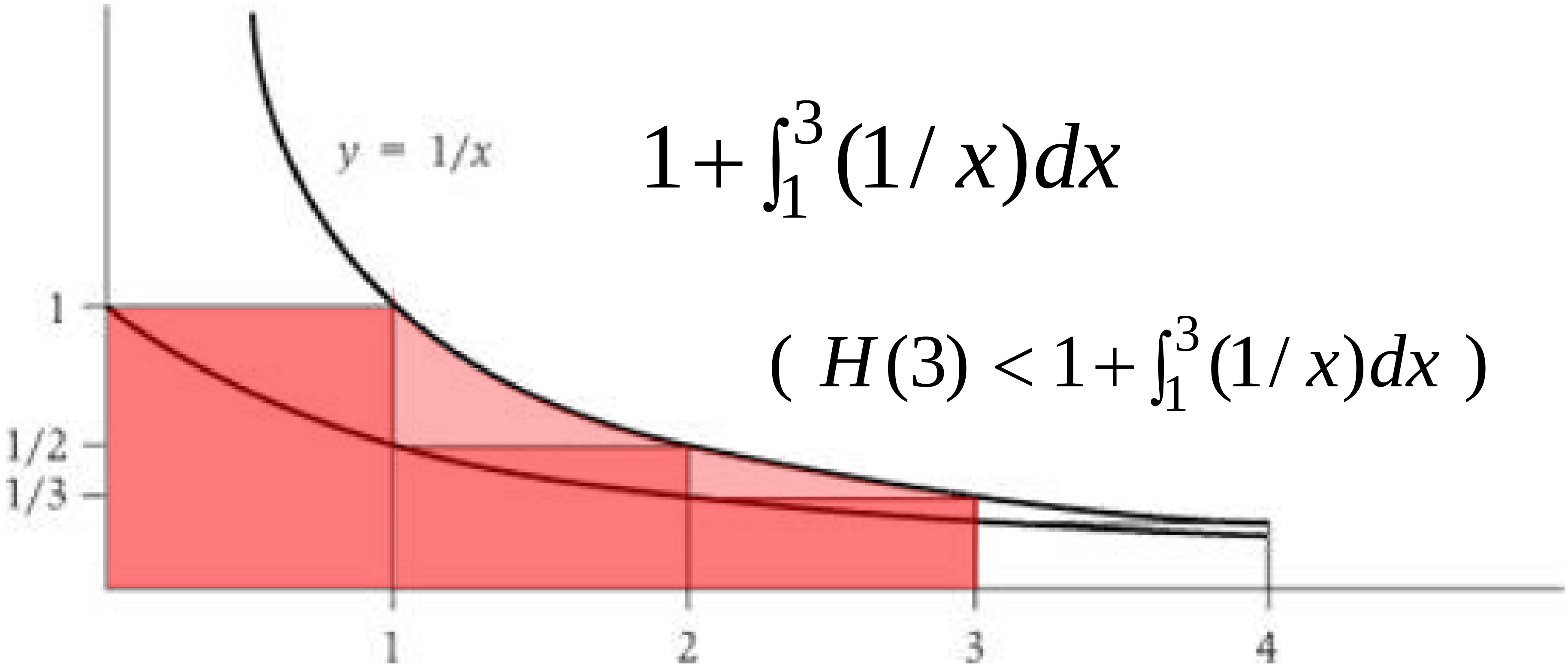
$$H(3) = \frac{1}{1} + \frac{1}{2} + \frac{1}{3}$$



***n* = 3**

Harmonic Function: $H(n) = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$

$$\int_1^{n+1} (1/x)\,dx = \ln(n+1) \;\le\; H(n) \;\le\; 1 + \int_1^{n} (1/x)\,dx = 1 + \ln(n)$$



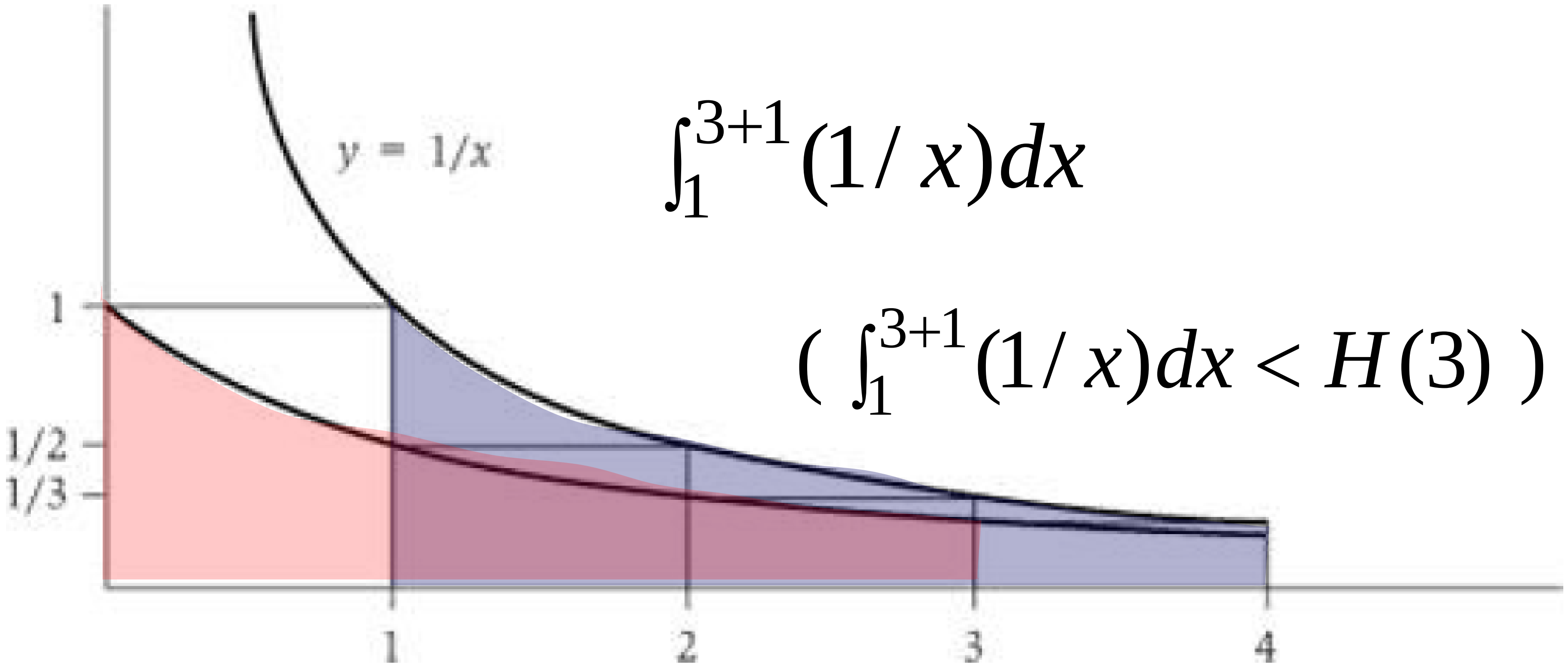
$$1 + \int_1^{3} (1/x)\,dx$$

$$\left(\; H(3) < 1 + \int_1^{3} (1/x)\,dx \;\right)$$

***n* = 3**

Harmonic Function: $H(n) = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$

$$\int_1^{n+1} (1/x)dx = \ln(n+1) \ \leq \ H(n) \ \leq \ 1 + \int_1^{n} (1/x)dx = 1 + \ln(n)$$



$$\int_1^{3+1} (1/x)dx$$

$$\left(\int_1^{3+1} (1/x)dx < H(3) \right)$$

***n* = 3**