

CSCI 2510 - Problem 1.4

Analysis (Sarah Eisenstat)

Much as we did in the approximation algorithm for set cover, we label all of the edges e_1, \dots, e_k in the order that they were covered. (That is, e_1 is one of the edges covered by the first vertex that we selected for the vertex cover, and e_k is one of the edges covered by the last vertex.)

For each of these edges, we let p_i be the “price” of the edge e_i — that is, the amount that it contributes to the total size of the vertex cover. For this problem, p_i is defined as the multiplicative inverse of the number of new edges that the most recently added vertex covers. Thus, when we sum up all of the p_i 's for the set of edges newly covered by a single vertex, we get a total of 1. Therefore, the sum of all p_i 's is equal to the total number of vertices.

So what sort of bounds can we place on p_i ? Well, we know that for the current sub-problem we're considering (before removing the vertex covering e_i), the maximum degree is equal to $\frac{1}{p_i}$. This means that the maximum number of edges covered by a single vertex is $\frac{1}{p_i}$, and so the number of vertices necessary to cover the remaining edges (at least $k - i + 1$ of them) is greater than or equal to $p_i(k - i + 1)$. The number of vertices in the optimum vertex cover, OPT , is also greater than or equal to that value. So we have:

$$\begin{aligned} OPT &\geq p_i(k - i + 1) \\ p_i &\leq \frac{OPT}{k - i + 1} \\ \sum_i p_i &\leq \sum_i \frac{OPT}{k - i + 1} \\ \sum_i p_i &\leq OPT \times H_n \\ \sum_i p_i &\leq OPT \times \log n \end{aligned}$$

Therefore, the size of the vertex cover (which, as we have previously established, is equal to the sum of all of the p_i 's) is less than or equal to $\log n^2 = 2 \log n$ times the size of the optimal vertex cover. Hence, we have an $O(\log n)$ approximation.

Tightness (Shay Mozes)

As a tight example, consider the following graph G on n nodes. There are $c \triangleq \frac{n}{\log n}$ special nodes, and $n - c$ regular nodes. There are $k \triangleq \log(c)$ groups of regular nodes, each with $\frac{c}{4}$ nodes. The remaining regular nodes are ungrouped and do not play part in the construction. For $i = 1, 2, \dots, k$, the nodes of the i th group are connected to $\frac{c}{2^{i-1}}$ of the special nodes, such that each special node is connected (up to ± 1) to the same number of nodes in each group and has the same degree (also, up to ± 1). Note that this is well defined since $2^{k-1} = \frac{c}{2}$, so the degree of a node in the i th group is smaller than the number of special nodes and greater than 2. Also note that the special nodes comprise a cover of G , so $OPT \leq c$.

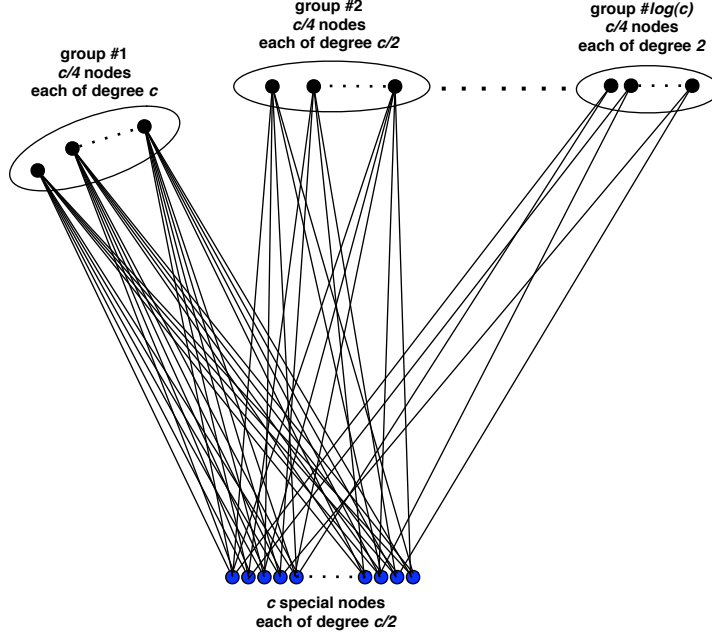


Figure 1: An illustration of the tight example. $c = \frac{n}{\log n}$

Lemma 1. For $i = 0, \dots, k - 1$, the degree of any special node at the end of the $(\frac{c}{4}i)$ -th iteration of the greedy algorithm is at most $\frac{c}{2^{i+1}}$. Furthermore, between the $(\frac{c}{4}(i - 1) + 1)$ -th and the $(\frac{c}{4}i)$ -th iterations, only nodes from group i are chosen by the algorithm, so group i is entirely consumed.

Proof. By induction on i . For the base case ($i = 0$), the total number of edges initially in the graph is:

$$\sum_{i=1}^k \frac{c}{4} \frac{c}{2^{i-1}} \leq \frac{c^2}{2}$$

Therefore, the degree of each special node is initially $\frac{c}{2}$, as required. The second part of the lemma is trivially empty for the base case.

Assume that the lemma holds for $i - 1$. Then at the end of the $(i - 1)\frac{c}{4}$ -th iteration, the degree of any special node is at most $\frac{c}{2^i}$. Furthermore, by the second part of the lemma, only nodes from group numbers strictly smaller than i were chosen so far, and all of these groups were entirely consumed. Since the degree of nodes in the i th group is $\frac{c}{2^{i-1}}$, the algorithm will keep choosing nodes from the i -th group for the next $\frac{c}{4}$ iterations, until that group is entirely consumed. This proves the second part of the lemma. To prove the first part, note that the total number of edges incident on group i is $\frac{c}{4} \frac{c}{2^{i-1}}$, so the degree of each of the special nodes decreases by the end of the $i\frac{c}{4}$ -th iteration by $\frac{c}{2^{i+1}}$ to:

$$\frac{c}{2^i} - \frac{c}{2^{i+1}} = \frac{c}{2^{i+1}},$$

as required. □

From the lemma it follows that all of the nodes in the k groups are chosen by the greedy algorithm. Hence, the size of the vertex cover returned by the greedy algorithm is

$$\frac{c}{4}k = \frac{n}{4 \log n} \log \left(\frac{n}{\log n} \right) = \frac{n}{4 \log n} (\log n - \log \log n) = O(n),$$

while we have seen that $OPT \leq \frac{n}{\log n}$, so the approximation factor is $\Omega(\log n)$, and this is indeed a tight example.