Recall that a vertex cover of an undirected graph \( G = (V, E) \) is a subset \( C \subseteq V \) such that if \( (u, v) \) is an edge of \( G \), then either \( u \in C \) or \( v \in C \) (or both). The size of a vertex cover is \( |C| \), the number of vertices in it. The vertex-cover problem is to find the optimal vertex cover, that is, the vertex cover of minimum size. As shown in section 31.5.2 of CLRS, this problem is NP-complete.

In section 35.1 of CLRS, the following simple-minded heuristic vertex cover algorithm is shown to approximate the optimal cover within a factor of 2:

1. Choose any edge.
2. Add its two vertices in the vertex cover.
3. Repeat after deleting the edges incident on either vertex of that edge.

Moreover, this ratio is achievable—in a graph with \( n \) independent edges on \( 2n \) vertices, the heuristic puts all \( 2n \) vertices in the cover, but only one of each pair is needed.

Exercise 35.1-3 suggests the following greedy heuristic:

1. Choose the vertex of highest degree.
2. Put it in the vertex cover and delete it and all edges incident on it from the graph.
3. Repeat until all the edges are deleted.

At first glance, this is a more sophisticated approach. However, its performance is worse than that of the simple-minded heuristic. Let \( C \) be the vertex cover thus found and let \( C^* \) be the optimal vertex cover. We will prove

**Theorem.** \( |C|/|C^*| \leq \lg n \).

Furthermore, we'll give a graph for which \( |C|/|C^*| = \Theta(\log n) \).

**Proof.** The greedy algorithm takes a vertex of maximum degree, \( v_1 \) in the original graph \( G = G_0 = (V_0, E_0) \), puts it into the vertex cover \( C \), and deletes it (and the edges incident on it) from \( G \), resulting in \( G_1 = (V_1, E_1) \). The process continues and after the \( i \)th vertex \( v_i \) is added to \( C \), the graph remaining is \( G_i = (V_i, E_i) \). Let’s examine the situation after the first \( |C^*| \) steps; call these first \( |C^*| \) steps the first “super iteration.” After this first super iteration the, \( |C^*| \) vertices \( v_1, v_2, \ldots, v_{|C^*|} \) have been added to the cover \( C \) and the graph remaining is \( G_{|C^*|} = (V_{|C^*|}, E_{|C^*|}) \). Note that if we got the optimal vertex cover, nothing would be left of the graph. However, we will prove in the lemma below that after a super iteration of \( |C^*| \) steps, at least half of the edges are gone and at most half remain. If we then continue with the next super iteration of \( |C^*| \) steps, beginning from graph \( G_{|C^*|} \) to graph \( G_{2|C^*|} \), we will have at most 1/4 of the edges left in the graph. Therefore, if we keep doing this, we see that there can be at most \( \lg n \) super iterations to cover all \( n \) edges. Each super iteration adds \( |C^*| \) vertices to the vertex cover \( G \), so that \( |C| \leq |C^*| \lg n \), and hence the approximation ratio is at most \( \lg n \). \( \square \)
Lemma. \( E_{|C^*|} \geq |E|/2. \)

Proof. The number of edges that disappeared in the first \(|C^*|\) iterations is:
\[
\deg(v_1) + \deg(v_2) + \cdots + \deg(v_{|C^*|}),
\]
where \( \deg(v_i) \) is the degree of vertex \( v_i \) in \( G_{i-1} \). Because \( C^* \) is a cover of \( G \), it is also a vertex cover of \( G_i \). Thus
\[
\sum_{x \in C^*} \deg(x) \geq |E_i|.
\]
Dividing both sides of this inequality by \(|C^*|\) tells us that the average degree in \( G_{i-1} \) is greater than or equal to \(|E_{i-1}|/|C^*|\). But the greedy choice means we pick the vertex of largest degree, which is at least as large as the average degree. That is, if we chose the vertex \( v_i \), we know that \( \deg(v_i) \geq |E_{i-1}|/|C^*| \). Because we always remove the chosen vertex from the graph, we know that
\[
\cdots \geq |E_{i-1}| \geq |E_i| \geq |E_{i+1}| \geq \cdots,
\]
so that \( |E_{|C^*|}| \) is the smallest of \( |E_0|, |E_1|, \ldots, |E_{|C^*|}| \). Hence,
\[
\sum_{i=1}^{|C^*|} \deg(v_i) \geq \sum_{i=1}^{|C^*|} |E_i|/|C^*| \geq |C^*| |E_{|C^*|}|/|C^*| = |E_{|C^*|}|. \tag{1}
\]
But \( |E_{|C^*|}| \) is the number of edges in the graph \( G_{|C^*|} \), so
\[
|E_{|C^*|}| = |E| - \sum_{i=1}^{|C^*|} \deg(v_i). \tag{2}
\]
Combining inequalities (1) and (2), we get
\[
|E_{|C^*|}| \geq |E| - |E_{|C^*|}|,
\]
and the lemma holds.

To see that the logarithmic ratio is achievable, consider the bipartite graph shown in the figure (on the next page) in which the left side has \( k \) vertices \( L = \{l_1, l_2, \ldots, l_k\} \) and the right side has vertices \( R = R_2 \cup R_3 \cup \cdots R_k \), where \( |R_i| = \lfloor k/i \rfloor \). Each vertex \( r \in R_i \) has \( i \) neighbors in \( L \); no two vertices in \( R_i \) share neighbors in \( L \). Thus each vertex in \( R_i \) has degree \( i \), but every vertex in \( L \) has degree at most \( k - 1 \) (at most one neighbor in each \( R_i \)). When the greedy heuristic is applied to this graph, it chooses a vertex cover consisting of all the vertices of \( R \). The number of vertices in \( R \) is
\[
|R| = \sum_{i=2}^k \lfloor k/i \rfloor \approx \sum_{i=2}^k k/i = kH_k \approx k \ln k.
\]
The number of vertices in the the graph is thus \( n = |L \cup R| = \Theta(k \log k) \), and there is a vertex cover of size \( k \) (the vertices of \( L \)). The ratio of the greedy heuristic’s vertex cover to the optimal vertex cover is thus at least \( \Theta(\log k) = \Theta(\log n) \).
Figure 1: Bipartite graph in which the greedy heuristic finds a vertex cover consisting of the of $\Theta(k \log k)$ vertices of $R$, but the $\Theta(k)$ vertices of $L$ are a smaller vertex cover.