Approximation algorithm for maximum independent set

This week, we analyze a greedy algorithm for the independent set problem. We again follow the section on independent sets in Chandra Chekuri's notes (2021).
Consider first a naive greedy algorithm for finding a maximal independent set: Let G be a simple graph.

Algorithm 1 Find Independent Set (Naive)

1: $S = \emptyset$
2: **while** $V \neq \emptyset$ **do**
3:     Add v to S, for some $v \in V(G)$
4:     Delete $v \cup N(v)$ from G
5: **end while**
6: Return S

Claim 1. *We have $|S| \geq \frac{n}{\Delta+1}$, where $n = |G|$ and $\Delta = \Delta(G)$, the maxmimum degree of G.*

Proof. We use a discharging method: place 1 charge on every vertex. Then, run the above algorithm. If a vertex is chosen to be in S, it loses its charge. If it is not chosen to be in S, then it is in the neighborhood of a vertex $v \in S$ in line 4 of the algorithm. It sends its charge to v. At the end of the algorithm, we have $|V \setminus S|$ charge total, which is all placed on vertices in S. Since each vertex in S received at most Δ charge, we have the inquality

$$n - |S| \leq |V \setminus S| \leq \Delta |S|,$$

and therefore

$$|S| \geq \frac{n}{\Delta + 1}.$$

□

Claim 2. *We have $|S| \geq \frac{\alpha(G)}{\Delta}$, where $n = |G|$ and $\alpha(G)$, the size of a largest independent set.*

Proof. Similarly to the previous claim, this time we place 1 charge on every vertex in some maximum independent set. Now, a vertex keeps its charge if it is chosen in the independent set, and otherwise discharges as before. This gives

$$\alpha(G) \leq \Delta |S|.$$

□

We may choose quite badly in the naive algorithm. For example, take a start graph $K_{1,n-1}$. If we are unlucky, we choose the middle vertex first, resulting in $|S| = 1$, while $\alpha(G) = n - 1$. This shows that the above bounds are sharp for this algorithm.
To improve this, we could randomize the choice at each iteration. Then, we could run this algorithm multiple times and keep the largest sets. This would work very well for solving the problem with the star. In class, we showed that

$$\mathbb{E}(|S|) = \frac{n}{d+1},$$

where d is the average degree of G, by using linearity of expectation.

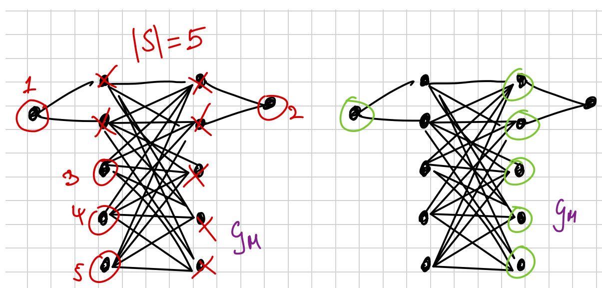
Consider instead the deterministic algorithm below, which prioritizes vertices of smallest degree. We depart from Chekuri's notes here. Exercise 4.1 in their notes asks us to show that we can find independent sets of size $\frac{n}{2(d+1)}$ using this greedy algorithm. We will show that it in fact achieves $\frac{n}{d+1}$.

Algorithm 2 Find Independent Set

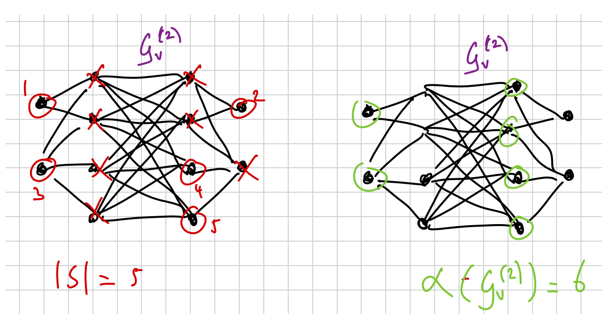
```
S = ∅
while G ≠ ∅ do
    Add v to S, such that d(v) = δ(G)
    Delete v ∪ N(v) from V
end while
Return S
```

In the case of the star graph example, this will always find the maximum independent set. Consider the following two examples which show that we do not always find an optimal set in general.

Example 3. *(Mazie O'Connor's Construction.) Consider the following graph G_M on 12 vertices. This graph consists of a $K_{5,5}$ with two additional vertices, as drawn below. We have $\alpha(G_M) = 6$, and $|S| = 5$ (always, by symmetry).*



Example 4. *(Vale Ferb's Construction.) Building on Mazie's idea, consider the following graph $G_V^{(k)}$ on $2^{k+1} + 2k$ vertices. Start with a $K_{2^k,2^k}$, labeled $v_1, \ldots, v_{2^k}$ and $w_1, \ldots, w_{2^k}$ (for the two partite sets). Then, add vertices $s_1, \ldots, s_k$, such that s_i has edges to $v_1, \ldots, v_{2^i}$, and similarly, $t_1, \ldots, t_k$, such that t_i has edges to $w_1, \ldots, w_{2^i}$. It is not difficult to see that $\alpha(G_V^{(k)}) = 2^k + k$. If we run the algorithm, without loss of generality, it will select $s_1, \ldots, s_k, t_1, \ldots, t_{k-1}, w_{2^{k-1}+1}, \ldots, w_{2^k}$. Therefore, $|S| = (1 + o(1))\frac{1}{2}\alpha(G_V^{(k)})$.*



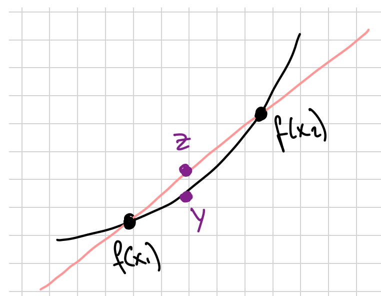
Now, we analyze the algorithm.

Claim 5. *The FIS algorithm returns an independent set S with $|S| \geq \frac{n}{d+1}$, where $d = d(G)$ is the average degree of G.*

We offer two proofs of this statement. Both use a version of Jensen's Inequality. We prove a baby version of Jensen's Inequality from scratch so that our first proof is complete. Let $f(x)$ be a real, convex function, and let $0, \leq, a, b \leq 1$ such that $a + b = 1$. Then

$$f(ax_1 + bx_2) \leq af(x_1) + bf(x_2).$$

Let $y = f(ax_1 + bx_2)$ and $z = af(x_1) + bf(x_2)$. Proof by picture:



Proof. Let G be an n-vertex graph and run the FIS algorithm. Let $S = \{v_1, \ldots, v_t\}$, with $d_i = d(v_i)$, $1 \leq i \leq t$. Each time a v_i is added to S, we delete d_i additional vertices. Since those vertices all had degree at least d_i, we also delete at least $\frac{1}{2}d_i(d_i + 1)$ edges. (Adding up the degrees of all deleted vertices and noting that an edge contributes to at most 2 degrees.) This implies that

$$\frac{1}{2}\sum_{i=1}^{t} d_i(d_i + 1) \leq |E| = \frac{nd}{2}. \tag{1}$$

Note that the left hand side represents the number of edges in some graph which consists of t cliques whose vertices add up to n. We claim that the fewest number of edges in such a graph is when the cliques are as close to equal in order as possible (differing by at most 1). Suppose tot he contrary that in a minimal such graph there are two cliques K_r and K_s with $s - r \geq 2$. If we remove a vertex from K_s we lose $s - 1$ edges, and when we add it to K_r we gain r edges, resulting in a graph on fewer edges; a contradiction. Let $n \equiv k \pmod{t}$. Then the number of edges in our minimal graph is

$$\begin{aligned}
&\left(k \cdot \binom{\lceil \frac{n}{t} \rceil}{2} + (t - k) \cdot \binom{\lfloor \frac{n}{t} \rfloor}{2}\right) \\
&= t \cdot \left(\frac{k}{t} \cdot \binom{\lceil \frac{n}{t} \rceil}{2} + \frac{t-k}{t} \cdot \binom{\lfloor \frac{n}{t} \rfloor}{2}\right) \\
&\geq t \cdot \frac{1}{2} \cdot \frac{n}{t}\left(\frac{n}{t} - 1\right) = \frac{1}{2}\left(\frac{n^2}{t} - n\right),
\end{aligned}$$

by baby Jensen and noting that $\binom{n}{2} = \frac{1}{2} \cdot n(n-1)$, the RHS of which is also a real, concex function. Combining this inequality with Inqeuality 1, we obtain that

$$t \geq \frac{n}{d+1}.$$

$\square$

Now, for the second version of the proof, we let X be a random variable. Then Jensen's Inequality states that
$$\mathbb{E}(X)^2 \leq \mathbb{E}(X^2).$$

Proof. We let d_S be the average degree of the vertices in S. If we think of X as a random variable that selects uniformly at random from $d_1, \ldots, d_t$. Then $\mathbb{E}(X) = d_S$. Then

$$\begin{aligned} d_S &= \frac{1}{t}\sum_{i=1}^{t} d_i(d_i+1) \\ &= \left(\frac{1}{t}\sum_{i=1}^{t} d_i^2\right) + \frac{1}{t}\sum_{i=1}^{t} d_i \\ &\geq d_S^2 + d_S = d_S(d_S+1), \end{aligned}$$

by Jensen's Inequality above. Now, we also note that during the algorithm, we removed n vertices, with $d_i + 1$ being removed at each step i. This implies that

$$\sum_{i=1}^{t} d_i + 1 = n,$$

and therefore
$$d_S + 1 = \frac{n}{t}.$$

Combining this with Inequality 1, we see that

$$\frac{n}{t}\left(\frac{n}{t}+1\right) = d_S(d_S+1) \leq \frac{nd}{t},$$

which rearranges to
$$t \geq \frac{n}{d+1}.$$

□