CIS 160

Recitation Guide - Week 8

Topics Covered: Trees, Independence

Problem 1:
Let T be a tree where the maximum degree is Δ. Prove that T has at least Δ leaves.

Solution:

We will use the (non-standard) notation $\lambda(T)$ to denote the number of leaves in a tree T. Thus, we can rewrite the claim as $\lambda(T) \geq \Delta$.

Direct Proof:

Let $v \in V$ have degree Δ in $T = (V, E)$. Consider the subgraph induced on the vertices $V \setminus \{v\}$. Each neighbor of v is in a distinct connected component in this graph, because we have destroyed the unique path between any two of v's neighbors in T when we removed v to induce the subgraph. Thus there are Δ connected components, each of which is a tree (as each connected component is still acyclic).

There are two possible cases for each connected component.

1. The connected component is a single node. Then this single node is a leaf adjacent to v in T.
2. The connected component has at least 2 nodes. Then, since the connected component is a tree, it must have at least 2 leaves (from Lecture 7T).

   Note that while it is true that one of these leaves in the induced subgraph could have been adjacent to v in T (therefore it wasn't a leaf in T), this can only be the case for at most one leaf in T. Otherwise there would have been a cycle between the two leaves and v in T.

In any case, each of the Δ connected components contains at least one leaf of T and hence T must have at least Δ leaves.

Maximal Path:

We first prove a lemma.

Lemma: All maximal paths in a tree must start and end with leaves.

Proof: Suppose for the sake of contradiction that at least one of the endpoints are not leaves. Let v be this vertex. Notice that $deg(v) \geq 2$, since it is not a leaf. As it is part of a maximal path, there must exist some vertex u that is a neighbor of v, but not the neighbor of v in the maximal path. However, this creates a cycle, as u must lie on the maximal path (or the path could be extended), which is a contradiction since trees do not have cycles. ■

Let $v \in V$ have degree Δ. For each $u_i, u_j \in N(v)$, let $P_{i,j}$ be a maximal path including $u_i - v - u_j$. Note that there must be at least $\binom{\Delta}{2}$ such paths, such that each pair of starting edges gives a different path. We know from the above lemma that any such path $P_{i,j}$ must terminate in two leaves. Lastly, note that since there is a unique path between any two vertices in a tree, every pair of leaves admits at most one maximal path. Therefore, if there were $\lambda(T) < \Delta$ leaves, we would end up with $\binom{\lambda}{2} < \binom{\Delta}{2}$ distinct maximal paths, a contradiction. We must then have $\lambda(T) \geq \Delta$.

Contradiction:

Assume that $\Delta \geq 2$, since the cases of $\Delta = 0$ and $\Delta = 1$ are clearly true. Suppose for the sake of contradiction that there are at most $\psi < \Delta$ leaves. Let $v \in V$ have degree Δ. Consider $S = \{u \in V \mid \{u, v\} \in E\}$. Note that S is the set of v's neighbors.

For all $u_i \in S$, there exists at least one path that starts with $\{v, u_i\}$ that ends with a leaf. We pick any such leaf for each edge $\{v, u_i\}$ and call the leaf l_i. Hence, by the Pigeonhole Principle, where the pigeons are the terminating leaves of each path and the holes are the ψ leaves available, we know that $\lceil \frac{\Delta}{\psi} \rceil \geq \lceil \frac{\Delta}{\Delta-1} \rceil = \lceil 1 + \frac{1}{\Delta-1} \rceil$ (since $\Delta \geq 2$) $= 2$ paths share the same terminating leaf, say ℓ_ω.

This is a contradiction, since the path between ℓ_ω and v are unique in a tree.

For each $u_i \in S$, let p_i be a maximal path starting from $v - u_i$. Note that there must be Δ such paths. We know from the lemma proven above that all such p_i must terminate in a leaf ℓ_i.

Induction on the number of vertices:

Let us prove this by induction on n, the number of vertices in the graph.

We formulate a proposition $P(n)$ which is: in a tree with n vertices and maximum degree Δ, the number of leaves in the tree is at least Δ.

Base Case: $n = 2$ and 3 There is only one possible tree when $n = 2$: $T = (V, E)$, $V = \{u, v\}$, $E = \{\{u, v\}\}$. Here $\Delta = 1$, and we have 2 leaves, so it checks out as required.

There is only one possible tree when $n = 3$: $T = (V, E)$, $V = \{u, v, w\}$, $E = \{\{u, v\}, \{v, w\}\}$. Here $\Delta = 2$, and we have 2 leaves, so it checks out as required.

We choose to show two base cases here to avoid a slightly unfortunate edge case in the Induction Step.

Induction Hypothesis: Assume that $P(k)$ is true, for some $k \in \mathbb{Z}^+, k \geq 2$.

Induction Step: Consider an arbitrary tree $T = (V, E)$ such that $|V| = k + 1$ and it has maximum degree Δ. Let $\ell \in V$ be an arbitrary leaf in T who has some neighbor a. Consider $T' = (V', E')$ where $V' = V \setminus \ell$ and $E' = E \setminus \{a, \ell\}$.

We know that $|V'| = k$ and T' is a tree (since removal of a leaf does not disconnect a tree, from Lecture 8T), so we can apply the Induction Hypothesis on T'.

Note that there are two cases here:

1. a was the only vertex of degree Δ in T.

   It must be the case then that a has degree $\Delta - 1$ in T' and is of maximum degree. The Induction Hypothesis gives us that T' must have at least $\Delta - 1$ leaves.

   Further note if a is a leaf in T', then $\Delta - 1 = 1$. $\Delta = 2$. Since a is the only vertex with degree Δ in T, then it must be the case that $n = 3$, and that is already shown to be true by the base case.

   Hence, going forward we will operate under the assumption that a is not a leaf.

   Adding ℓ back to T' to reconstruct T increases the number of leaves by one (since v is not a leaf), so we have that T has at least Δ leaves.

2. There is some vertex in T' that has degree Δ.

   By the Induction Hypothesis, we have that T' must have Δ leaves.

   There are two more cases here:

   (a) a is a leaf in T'

   In this case, the addition of ℓ back to T' does not change the number of leaves, which means we have at least Δ leaves in T, as desired.

   (b) a is not a leaf in T'

   In this case, the addition of ℓ to T' increases the number of leaves by 1, which means we have at least $\Delta + 1$ leaves in T, which proves our claim.

Induction on the number of edges:

You can do a similar procedure to the induction on the number of vertices in order to perform induction on the number of edges.

Strong Induction on the number of edges:

Let us prove this by induction on m, the number of edges in the graph.

We formulate a proposition $P(m)$ which is: in a tree with m edges and maximum degree Δ, the number of leaves in the tree is at least Δ.

Base Case: $m = 1$ and 2 There is only one possible tree when $m = 1$: $T = (V, E)$, $V = \{u, v\}$, $E = \{\{u, v\}\}$. Here $\Delta = 1$, and we have 2 leaves, so it checks out as required.

There is only one possible tree when $m = 2$: $T = (V, E)$, $V = \{u, v, w\}$, $E = \{\{u, v\}, \{v, w\}\}$. Here $\Delta = 2$, and we have 2 leaves, so it checks out as required.

We choose to show two base cases here to avoid a slightly unfortunate edge case in the Induction Step.

Induction Hypothesis: Assume that $P(j)$ is true, for $j \in \mathbb{Z}^+, 1 \leq j \leq k$, for some $k \in \mathbb{Z}^+, k \geq 2$.

Induction Step: Let T be a tree with $k+1$ edges and with a maximum degree Δ. Let v be a vertex with degree Δ, and u be an arbitrary neighbor of v. Let us consider $G' = (V', E')$, where $V' = V$, $E' = E \setminus \{\{u, v\}\}$. Note that G' must have had two connected components C_1 and C_2, which are both trees when a subgraph is induced on each of them. Let C_1 be the component with v, and let C_2 be the component with u.

There are two cases here:

1. There is another vertex in C_1 that has degree Δ

   From the induction hypothesis, we have that there must be Δ leaves in C_1. Let us reconstruct T from G'.

   There are two cases here:

   (a) $|C_2| = 1$

   In this case, if v is a leaf in G', then the addition of $\{u, v\}$ will not change the number of leaves. Therefore we have that T must have at least Δ leaves. If v is not a leaf, then the addition of $\{u, v\}$ will add an additional leaf, so we have that T must have at least $\Delta + 1$ leaves.

(b) $|C_2| \geq 2$

In this case, C_2 must have two leaves (from lecture 8T). Hence there are at least $\Delta + 2$ leaves in G'. Notice that the addition of the edge $\{u, v\}$ can decrease the number of leaves by up to 2 (if u and v were both leaves in G'). Hence we have that T has at least Δ leaves, as required.

2. v is the only vertex with degree Δ in T.

Hence, $\Delta(C_1) = \Delta - 1$. From the induction hypothesis, we know that C_1 must have $\Delta - 1$ leaves. We further note that if v is a leaf in G', $\Delta - 1 = 1$, $\Delta = 2$. Since v is the only vertex with degree 2, it must be that $m = 2$, and we have already shown the validity of this in the base case. We will therefore operate now under the assumption that v is not a leaf.

There are two cases here:

(a) $|C_2| = 1$

Since v is not a leaf, then the addition of $\{u, v\}$ will add an additional leaf, so we have that T must have at least Δ leaves.

(b) $|C_2| \geq 2$

In this case, C_2 must have two leaves. Hence there are at least $\Delta + 1$ leaves in G'. Notice that the addition of the edge $\{u, v\}$ can decrease the number of leaves by up to 1 (if u is a leaf in G'). Hence we have that T has at least Δ leaves, as required.

Using inequalities:

We know that a tree with n vertices must have $n - 1$ edges. Since the sum of the degrees of all the vertices in a graph must be twice the number of edges, we know that the total of all degrees in the tree must be $2n - 2$.

Let us consider the following partitioning of the vertices in V. Let $A = \{v \in V \mid deg(v) = \Delta\}$, $B = \{v \in V \mid 1 < deg(v) < \Delta\}$, and $C = \{v \in V \mid deg(v) = 1\}$. Note that $V = A \cup B \cup C$ and $A \cap B = \emptyset, A \cap C = \emptyset, B \cap C = \emptyset$. Note that C is the set of leaves.

$$\begin{aligned}
2n - 2 &= \sum_{v \in V} deg(v) \\
&= \sum_{v \in A} deg(v) + \sum_{v \in B} deg(v) + \sum_{v \in C} deg(v) \\
&= \Delta \cdot |A| + \sum_{v \in B} deg(v) + |C| \\
&\geq \Delta \cdot |A| + |C| + 2 \cdot |B| \\
&= \Delta \cdot |A| + |C| + 2 \cdot (n - |A| - |C|) \\
&= (\Delta - 2) \cdot |A| - |C| + 2n \\
&\geq (\Delta - 2) - |C| + 2n
\end{aligned}$$

Hence we have established that $2n - 2 \geq (\Delta - 2) - |C| + 2n$. Further, we have that:

$$
\begin{aligned}
2n - 2 &\geq (\Delta - 2) - |C| + 2n \\
-2 &\geq \Delta - 2 - |C| \\
|C| &\geq \Delta
\end{aligned}
$$

Hence we have that the number of leaves is at least Δ.

Problem 2:
Suppose E and F are independent events and $\Pr[E] > 0$. Let $\overline{F}$ denote the complement of F. Are E and $\overline{F}$ independent? Prove your answer.

Solution:

As E and F are independent events,

$$\Pr[E \cap F] = \Pr[E] \cdot \Pr[F]$$

According to total probability theorem,

$$\begin{aligned}
\Pr[E] &= \Pr[E \cap F] + \Pr[E \cap \overline{F}] \\
&= \Pr[E] \cdot \Pr[F] + \Pr[E \cap \overline{F}] && \text{(given that E and F are independent)} \\
\Pr[E \cap \overline{F}] &= \Pr[E] - \Pr[E] \cdot \Pr[F] \\
&= \Pr[E] \cdot (1 - \Pr[F]) \\
&= \Pr[E] \cdot \Pr[\overline{F}]
\end{aligned}$$

Therefore, if E and F are independent, then, E and $\overline{F}$ are independent.

Problem 3:
We have three wooden buckets, T_A, T_B, T_C and we throw $n \geq 3$ metal keys in them. The key throws are mutually independent and each key is equally likely to land in each of the three buckets.

(a) Let A be the event that after all keys are thrown bucket T_A has at least one key in it and similarly associate an event B with T_B. Are A and B independent? Justify your answer.

(b) Compute the probability that after all keys are thrown, each of the three buckets has at least one key in it. Justify your answer.

Solution:

Define $\Omega = \{(x_1, x_2, ..., x_n) \mid x_i \in \{A, B, C\}\}$. That is, the sample space is the set of all n-tuples where the ith element in the tuple represents the bucket that key i landed in.

(a) For $i = 1, \ldots, n$ let A_i be the event that key i is thrown in bucket T_A. We have $\Pr[A_i] = \frac{1}{3}$. Clearly $A = A_1 \cup \cdots \cup A_n$ and since the events $A_1, \ldots, A_n$ are mutually independent we can compute:

$$\Pr[A] = \Pr[A_1 \cup \cdots \cup A_n] \;=\; 1 \;-\; \prod_{i=1}^{n} (1 - \Pr[A_i]) \;=\; 1 - \left(1 - \frac{1}{3}\right)^n \;=\; 1 - \left(\frac{2}{3}\right)^n$$

Similarly, $\Pr[B] = 1 - \left(\frac{2}{3}\right)^n$. To check independence we also need $\Pr[A \cap B]$.

Upon reflection, we notice that there is one aspect of the problem that we have not used yet: the keys get thrown *only* in T_A, T_B and T_C. Thus, $\bar{A} \cap \bar{B}$, which means that both T_A and T_B are empty after all keys are thrown, is the same as the event "all keys get thrown in T_C" and therefore, by mutual independence, has probability $\left(\frac{1}{3}\right)^n$, as each key has a $\frac{1}{3}$ probability of being thrown into T_C. Now we can compute, using properties of probability and De Morgan's Laws:

$$\begin{aligned}
\Pr[A \cap B] &= \Pr[A] + \Pr(B] - \Pr[A \cup B] \\
&= \Pr[A] + \Pr[B] - (1 - \Pr[\bar{A} \cap \bar{B}]) \\
&= 1 - \left(\frac{2}{3}\right)^n + 1 - \left(\frac{2}{3}\right)^n - \left(1 - \left(\frac{1}{3}\right)^n\right) \\
&= 1 - 2\left(\frac{2}{3}\right)^n + \left(\frac{1}{3}\right)^n
\end{aligned}$$

But we also know that:

$$\Pr[A] \cdot \Pr[B] = \left(1 - \left(\frac{2}{3}\right)^n\right)\left(1 - \left(\frac{2}{3}\right)^n\right) = 1 - 2\left(\frac{2}{3}\right)^n + \left(\frac{4}{9}\right)^n$$

Since $\frac{1}{3} \neq \frac{4}{9}$ it follows that $\Pr[A] \cdot \Pr[B] \neq \Pr[A \cap B]$ hence A and B are *not* independent.

(b) We continue with the notation introduced in part (a) and we also define C to be the event "T_C is not empty after all keys are thrown. This part asks for $\Pr[A \cap B \cap C]$. We are tempted to multiply probabilities but we do not know if A, B, C are mutually independent. In fact, in part

(a) we saw that $A \not\perp B$. Although it is still possible that $\Pr[A \cap B \cap C] = \Pr[A] \cdot \Pr[B] \cdot \Pr[C]$ there is no reason to hope for this here (and in fact we shall see that it does not hold).

Instead, we will use the Principle of Inclusion-Exclusion for three events:

$$\Pr[A \cup B \cup C] = \Pr[A] + \Pr[B] + \Pr[C] - \Pr[A \cap B] - \Pr[B \cap C] - \Pr[C \cap A] + \Pr[A \cap B \cap C]$$

Since we have at least one key, at least one of the buckets ends up non-empty. Hence $A \cup B \cup C = \Omega$, meaning $A \cup B \cup C$ consists of all the outcomes and has probability 1. From part (a) we have:

$$\Pr[A] = \Pr[B] = \Pr[C] = 1 - \left(\frac{2}{3}\right)^n$$

$$\Pr[A \cap B] = \Pr[B \cap C] = \Pr[C \cap A] = 1 - 2\left(\frac{2}{3}\right)^n + \left(\frac{1}{3}\right)^n$$

We plug in and obtain

$$\begin{aligned}
\Pr[A \cap B \cap C] &= \Pr[A \cup B \cup C] - \Pr[A] - \Pr[B] - \Pr[C] + \Pr[A \cap B] + \Pr[B \cap C] + \Pr[C \cap A] \\
&= 1 - 3\left(1 - \left(\frac{2}{3}\right)^n\right) + 3\left(1 - 2\left(\frac{2}{3}\right)^n + \left(\frac{1}{3}\right)^n\right) \\
&= 1 - 3\left(\frac{2}{3}\right)^n + 3\left(\frac{1}{3}\right)^n
\end{aligned}$$