Graphs: Connectivity, Trees[1]

- **Perambulations in Graphs.** We introduce a lot of definitions involving alternating sequence of vertices and edges. These are key definitions so make sure you understand them. Throughout below we fix a graph $G = (V, E)$.
  - A ***walk*** w in G is an *alternating sequence* of vertices and edges
$$w = (v_0, e_1, v_1, e_2, v_2, \ldots, e_k, v_k)$$
    such that the ith edge $e_i = (v_{i-1}, v_i)$ for $1 \le i \le k$.
    Intuitively, imagine starting at vertex v_0, using the edge e_1 to go to the adjacent vertex v_1, and then using e_2 to go to the adjacent (to v_1) vertex v_2, and so on and so forth till we reach v_k.
    Note both the edges and vertices could repeat themselves. That is e_i could be the same as e_j for $j \neq i$. In fact, e_{i+1} could be the same as e_i; this would mean going from one endpoint of e_i to the other and immediately returning back.
    The walk above is said to ***start*** at v_0 and ***end*** at v_k. The node v_0 is often called the ***source/origin*** and the node v_k is often called the ***sink/destination***. If there is a walk as described above, then we often say "there is a walk from v_0 to v_k."
    A walk is of ***length*** k if there are k edges in the sequence. Note that since repetition of both vertices and edges are allowed, walks could go on for ever.
  - A ***trail*** t in G is a walk with no edges repeating. That is, a trail is also an alternating sequence of vertices and edges
$$t = (v_0, e_1, v_1, e_2, v_2, \ldots, e_k, v_k) \quad \text{where the } e_i\text{'s are distinct}$$
    Note that a trail *could repeat vertices*. For instance, if the graph was
    $G = (\{a, b, c, d, e\}, \{(a, b), (b, c), (c, d), (d, b), (b, e)\})$, then the following is a valid trail. The vertex b is repeated.
$$t = (a,\ (a, b),\ b,\ (b, c),\ c,\ (c, d),\ d,\ (d, b),\ b,\ (b, e),\ e)$$
    Also note that a trail cannot be arbitrarily long. A trail's length is at most $|E|$.
  - A ***path*** p in a graph G is a walk with no *vertices repeated*. Note that a path is always a trail. In fact, a path is a trail with no vertices repeating. Oftentimes, for describing paths, the alternating edges are dropped. So for instance
$$p = (v_0, v_1, \ldots, v_k) \text{ actually stands for } (v_0,\ (v_0, v_1),\ v_1,\ (v_1, v_2),\ v_2, \cdots, (v_{k-1}, v_k),\ v_k)$$
  - A ***closed walk*** is a walk whose origin and destination are the same vertex. If $e = (u, v)$ is an edge in G, then the following is a closed walk of length 2
$$w = (u, e, v, e, u)$$

[1] *Lecture notes by Deeparnab Chakrabarty. Last modified : 28th Aug, 2021*
These have not gone through scrutiny and may contain errors. If you find any, or have any other comments, please email me at deeparnab@dartmouth.edu. Highly appreciated!

A closed walk must be of length at least 2.

Note that given a closed walk, we can choose any $v_i \in w$ to be the source *and* the destination using the same vertices and edges of the closed walk. That is, given a closed walk

$$w = (v_0, e_1, v_1, e_2, v_2, \ldots, e_k, v_k) \quad \text{with } v_k = v_0$$

and an arbitrary vertex $v_i \in w$ with $1 \le i < k$, we can have another closed walk

$$w' = (v_i, e_{i+1}, v_{i+1}, \ldots, e_k, v_k = v_0, e_1, v_1, e_2, v_2, \ldots, e_i, v_i)$$

Note w' is a closed walk whose source and destination are v_i.

- A ***circuit*** is a closed trail of length at least 1. That is, it is a trail whose origin and destination are the same vertex, and contains at least one edge. The latter constraint disallows a singleton node from being defined as a circuit. Indeed, a circuit must have at least 3 edges – do you see this?
- A ***cycle*** is a circuit with no vertex other than the source and destination repeating. Thus, a cycle is a path followed by an edge from the destination of the path to the origin, and then the origin node.

Theorem 1. Let $G = (V, E)$ be a graph and u and v be two distinct vertices in $V(G)$. If there is a *finite walk* from u to v in G, then there is a *path* from u to v.

Proof. In the UGP, you see a way to prove the above by induction. However, for this theorem, going via the "minimal counter example" idea is way better. Goes like this.

Let W be the set of all walks from u to v of finite length. By the premise of this theorem, we know that this is set is non-empty. Pick a walk $w \in W$ from u to v of the ***smallest length***. We claim that this walk must be a path which would prove the theorem.

Suppose, for the sake of contradiction, w is not a path. That is,

$$w = (x_0 := u, e_1, x_1, \ldots, x_i, e_{i+1}, x_{i+1}, \ldots, x_j, e_{j+1}, x_{j+1}, \ldots, x_{k-1}, e_k, x_k := v)$$

but two vertices, say x_i and x_j with $i < j$ and both $0 \le i, j \le k$, are the same. Note that the length of the walk is k. Also note, that j cannot be $i + 1$ since this is true for every walk. Therefore, since $j > i$, it must be $j \ge i + 2$. Then, consider the walk

$$w' = (x_0 := u, e_1, x_1, \ldots, x_i, e_{j+1}, x_{j+1}, \ldots, e_k, x_k)$$

The length of this walk w' is $k - (j - i + 1) \le k - 1$. This walk w' is a smaller length walk than w. But this contradicts the choice of w. Thus, our supposition must be wrong. Therefore, w is a path. □

Theorem 2. Let $G = (V, E)$ and u be an arbitrary vertex in $V(G)$ and e be an arbitrary edge in $E(G)$. If there is a *circuit* in G containing u, then there is a *cycle* in G containing u. If there is a *circuit* in G containing e, then there is a *cycle* in G containing e.

Proof. Let me prove the first statement, and leave the second as an exercise (it can be found in the UGP).

We are given that there is at least one circuit containing u. Among all these circuits, let C be a circuit containing u of the smallest length (once again, there could be many, and we pick one arbitrarily). Let

$$C = (u = u_0, e_1, u_2, \ldots, u_k = u)$$

We claim that C is a cycle.

Suppose, for contradiction's sake C is not a cycle. Then, C has a repeating vertex v. If u_0 is this repeated vertex, then there must exists $0 < j < k$ such that $u_j = u_0 = v$. Now note that C can be "broken" into two circuits as follows

$$C_1 = (u = u_0, e_1, u_1, \ldots, e_j, u_j = u) \quad \text{and} \quad C_2 = (u, e_{j+1}, u_{j+1}, \ldots, e_k, u_k = u)$$

Both of these circuits are smaller in length than C (one is of length j and the other $k - j$). At least one of them contains the vertex u, and thus would lead to a smaller circuit containing the vertex u. This is a contradiction. Therefore, u_0 is not the repeated vertex.

Therefore, the repeated vertex is some $u_i = u_j$ for $0 < i < j < k$. That is,

$$C = (u = u_0, e_1, u_1, \cdots e_i, u_i, e_{i+1} \ldots, e_j, u_j, e_{j+1}, u_{j+1}, \cdots, e_k, u_k = u)$$

where $u_i = u_j = v$ and $i < j$. Again, as argued in the previous theorem, $j > i + 1$. Now consider,

$$C' = (u = u_0, e_1, u_1, \cdots, e_i, u_i, e_{j+1}, u_{j+1}, \cdots, e_k, u_k = u)$$

This is a circuit containing u which contains $k - (j - i + 1) < k$ edges. This contradicts the choice of C. Therefore, C must be a cycle. $\square$

- **Connectivity in Graphs**

  In a graph $G = (V, E)$, we say a vertex v is ***reachable*** from u if there is a path starting from v and ending at u.

  A graph $G = (V, E)$ is ***connected*** if any vertex u is reachable from another vertex v. A graph is ***disconnected*** otherwise.

  Given any graph $G = (V, E)$, we can partition it into ***connected components***. That is, $V = V_1 \cup V_2 \cup \cdots \cup V_k$ where (a) any two vertices in the same V_i are reachable from one another, and (b) a vertex $u \in V_i$ is *not* reachable from any vertex $v \in V_j$ if $i \neq j$.

  Given any graph $G = (V, E)$ and a vertex $u \in V$, the set of vertices $S_u \subseteq V$ which are *reachable* from u is the connected componenet of G which contains u.

- **Trees.**

  A graph $G = (V, E)$ is a ***forest*** if it doesn't contain any cycles.

Theorem 3 (Low Degree Theorem)**.**
Let $G = (V, E)$ be a forest. There must exist a vertex $v \in V$ with $\deg_G(v) \leq 1$.

Proof. Suppose, for the sake of contradiction, $\deg_G(v) \geq 2$ for all vertices $v \in V$. Therefore, $E \neq \emptyset$ (among other things, the handshake lemma shows this). Now, consider the *longest path* p in G. Since $E \neq \emptyset$, there is at least one path in G, and thus p is well-defined. Let

$$p = (x_1, x_2, \dots, x_{k-1}, x_k)$$

Now, by our supposition, $\deg_G(x_1) > 1$. Therefore, there exists a vertex $v \neq x_2$ such that (v, x_1) is an edge. Note: v cannot be x_i for any $3 \leq i \leq k$. For if that was the case, then $(x_1, x_2, \dots, x_{i-1}, x_i = v, x_1)$ would be a cycle. And G has no cycle (being a forest). Therefore, since $v \neq x_1$ and $v \neq x_2$, $v \notin p$. But then $(v, x_1, x_2, \dots, x_k)$ is a *longer path* than p. This contradicts the choice of p being the longest path.

□

A forest $G = (V, E)$ is a ***tree*** if it is connected. That is, a tree $G = (V, E)$ is a connected graph which doesn't contain any cycles.

Theorem 4. Let $G = (V, E)$ be a forest. Then each connected component of G induces a tree.

Proof. Let $V_1, \dots, V_k$ be the connected components of G. Each $G[V_i]$ is connected by definition. If G doesn't contain a cycle, then any subgraph also doesn't contain a cycle. Thus, $G[V_i]$ contains no cycle. Thus $G[V_i]$ is a tree. □

There are many equivalent ways to think about trees. We prove some here, and some are left as exercises in the UGP and the Pset.

Theorem 5 (Leaves of a tree)**.** Let $G = (V, E)$ be a tree with $|V| \geq 2$. There must exist a vertex $v \in V(G)$ with $\deg_G(v) = 1$. All such degree 1 vertices are called ***leaves***.

Proof. This is really a corollary to Theorem 3. Since G is connected and $|V| \geq 2$, every vertex $v \in V$ must have $\deg_G(v) \geq 1$. If $\deg_G(v) = 0$, then v would be isolated and not connected to any other vertex in G. Then this theorem follows from Theorem 3. □

Theorem 6 (Number of edges in a tree.)**.** Let $G = (V, E)$ be a tree. Then $|E| = |V| - 1$.

Proof. We prove this by induction (using the minimal counter example idea). Suppose the above statement is not true. Let G be a counterexample graph with the *smallest number of vertices*. Note, $|V(G)| > 2$ since the only tree on 2 vertices is a single edge, and that satisfies the statement of the theorem.

Since G is a tree, by Theorem 5, there exists a leaf vertex $v \in V(G)$ with $\deg_G(v) = 1$. Let (v, u) be the unique edge in $E(G)$. Consider the graph $G' = G - v$. Note, $E(G') = E(G) - (u, v)$. Thus, $|E(G)| = |E(G')| + 1$. Also, $|V(G)| = |V(G')| + 1$.

We now claim that G' is a tree. Since we obtained G' from G by deleting a vertex, and since G had no cycles, G' has no cycles. Is G' connected? Pick any two vertices $x, y \in V(G')$. Since G was connected, there is a path $p = (x = x_0, x_1, \ldots, x_k = y)$ in G from x to y. However, none of the internal nodes x_i for $1 \leq i \leq k-1$ can be v; this is because internal nodes have $\deg_G(x_i) \geq 2$. Therefore, this path p still exists in G', implying x and y are connected in G'. Since x and y were an arbitrary pair of vertices in $V(G')$, we can claim that G' is connected.

That is, G' is connected and acyclic — G' is a tree. Since $|V(G')| < |V(G)|$, and G was the smallest counterexample tree, we get $|E(G')| = |V(G')| - 1$. But this implies $|E(G)| = |V(G)| - 1$. Which contradicts G was a counterexample. □

Theorem 7. Let $G = (V, E)$ be a forest with k connected components. Then $|E(G)| = |V(G)| - k$.

Proof. Let $V_1, \ldots, V_k$ be the connected components of $G = (V, E)$. We know that each $G_i := G[V_i]$ is a tree. Thus, $|E(G_i)| = |V(G_i)| - 1$. Thus, $|E(G)| = \sum_{i=1}^{k} |E(G_i)| = \left(\sum_{i=1}^{k} |V(G_i)|\right) - k = |V(G)| - k$. □

The next theorem shows that for connected graphs, the relation between the number of vertices and edges is sufficient for "tree-ness".

Theorem 8. Let G be a connected graph with $|E(G)| = |V(G)| - 1$. Then G is a tree.

Proof. The proof is similar to the above proof. We first assume $|V(G)| \geq 2$ for otherwise, G would be a singleton vertex and vacuously a tree.

The main observation is the following.

> If $|E(G)| = |V(G)| - 1$ for some connected graph, then there must exist $v \in V(G)$ with $\deg_G(v) = 1$.

This follows from the handshake lemma. Since G is connected and $|V(G)| \geq 2$, we get $\deg_G(v) \geq 1$ for all v. If there was no degree 1 vertex, then we would get $\deg_G(v) \geq 2$ for all v, and the handshake lemma would imply $2|E(G)| \geq 2|V(G)|$. This contradicts $|E(G)| = |V(G)| - 1$.

Now suppose G is the smallest counterexample to the above theorem. Let v be any vertex in G with $\deg_G(v) = 1$. Let $(v, u) \in E(G)$ be the unique edge incident on v. We consider the graph $G' = G - v$. As in the previous theorem, one can argue G' is connected (I leave this to the reader – but please try to write this proof without looking above). Since $|E(G')| = |V(G')| - 1$, and G was the smallest counterexample, we get G' is a tree. But $G = G' + (v, u)$ for some $v \notin V(G')$. This introduces no cycles, and v is connected to any other $x \in V(G)$ because u is connected to any other $x \in V(G)$. This implies G is a tree. □

Exercise: *Show that the connectedness is needed. That is, describe a graph with G with $|E(G)| = |V(G)| - 1$ which is* **not** *a tree.*

Another proof of Theorem 8. Let me also add a proof via the usual induction route (predicates and all). Here goes.

Define $P(n)$ to be the predicate which takes the value true if *for all* connected graphs $G = (V, E)$ on n vertices with $|E(G)| = |V(G)| - 1$, the graph G is a tree. We want to prove $\forall n \in \mathbb{N} : P(n)$ is true. We proceed via induction.

Base Case. $P(1)$ is true. Vacuously/Obviously. Any graph on 1 vertex is a tree.

Inductive Case. Fix any natural number $k \geq 1$. Assume $P(k)$ is true. We desire to prove $P(k+1)$ is true. That is, we desire to prove

> Given ***any*** graph $G = (V, E)$ with $|V| = k+1$, and G is connected, and $|E| = |V| - 1$, then G is a tree.

To this end, *fix* and *arbitrary* connected graph G which has $k+1$ vertices and k edges. We need to show this is a tree.

Ok, we already know G is connected. So we need to prove G has no cycles.

Observation 1: G is connected, and $|V(G)| \geq 2$, so $\deg_G(v) \geq 1$. There cannot be 0 degree vertices.

Key Observation: Handshake lemma gives us $\sum_{v \in V(G)} \deg_G(v) = 2|E(G)| = 2(|V(G)| - 1) = 2|V(G)| - 2$. This implies *some* vertex $v \in V(G)$ ***must*** have $\deg_G(v) = 1$. Why? If not, then all vertices have $\deg_G(v) \geq 2$, which in turn would imply $\sum_{v \in V(G)} \deg_G(v) \geq 2|V(G)|$. But the sum is $< 2|V(G)|$.

Now, consider this vertex v and the edge (v, u) in the graph G. *Construct* a new graph $G' := G - v$. That is, delete vertex v and the edge (u, v). Two things:

- $|E(G')| = |E(G)| - 1$, $|V(G')| = |V(G)| - 1$, and therefore
$$|E(G')| = |V(G')| - 1$$
- G' is connected. This prood is similar to the previous theorem and I am leaving this as an exercise.

Therefore, since $P(k)$ is true, and since $|V(G')| = k$, G' is connected, and $|E(G')| = |V(G')| - 1$, we get by the induction hypothesis, G' is a tree. In particular, G' has no cycles.

Remember what we had to show? We had to show G had no cycles. Well suppose it did. Every vertex in this cycle would have degree ≥ 2 since they lie on a cycle. In particular, v is not on this cycle, and neither is the edge (u, v). That is, this cycle is also present in G'. But G', we just established, has no cycles. And so, we can assert G has no cycles.

By induction, we have $\forall n \in \mathbb{N} : P(n)$ is true, which is precisely what we wanted to prove. $\square$