Theorem. Let G be a graph with n nodes and e edges. Then the following are equivalent:

1. G is a tree (connected and acyclic).
2. Every two nodes of G are joined by a unique path.
3. G is connected and $n = e + 1$.
4. G is acyclic and $n = e + 1$.
5. G is acyclic and if any two nonadjacent points are joined by a line, the resulting graph has exactly one cycle.

Proof. We need to show that $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5 \Rightarrow 1$.

$1 \Rightarrow 2$: *If G is a tree then every two nodes are joined by a unique path.* Suppose that there are two paths P_1 and P_2 between node u and node v. Tracing the two paths simultaneously from u to v, let w be the first point that is on both paths, but for which the successor points are on different paths. Also, let x be the next point after w that is is on both paths. Then the paths between w and x on P_1 and P_2 together make a cycle. But this can't happen if G is acyclic.

$2 \Rightarrow 3$: *If every two nodes of G are joined by a unique path, then G is connected and $n = e + 1$.* G is connected since any two nodes are joined by a path. To show $n = e + 1$, we use induction. Assume it's true for less than n points. Removing any edge from G breaks G into two components, since paths are unique. Suppose the sizes are n_1 and n_2, with $n_1 + n_2 = n$. By the induction hypothesis, $n_1 = e_1 + 1$ and $n_2 = e_2 + 1$; but then $n = n_1 + n_2 = (e_1 + 1) + (e_2 + 1) = (e_1 + e_2) + 2 = e - 1 + 2 = e + 1$.

$3 \Rightarrow 4$: *If G is connected and $n = e + 1$ then G is acyclic.* Suppose G has a cycle of length k. Then there are k points and k edges on this cycle. Since G is connected, for each node v not on the cycle, there is a shortest path from v to a node on the cycle. Each such path contains an edge e_v not on any other (since they are shortest paths). Thus, the number of edges is at least $e \geq (n - k) + k = n$, which contradicts the assumption $n = e + 1$.

$4 \Rightarrow 5$: *If G is acyclic and $n = e+1$ then if any two nonadjacent points are joined by a line, the resulting graph has exactly one cycle.* Since G doesn't have cycles, each component of G is a tree. Suppose there are k components. Thus, $n_i = e_i + 1$ if the i-th component has e_i edges and n_i nodes, and therefore $n = e + k$. It follows that $k = 1$, so G is in fact connected, and therefore a tree. For any pair of disconnected nodes u and v, there is a unique path between them. Adding the the line (u, v) thus results in a single cycle.

$5 \Rightarrow 1$: *If G is acyclic joining any nonadjacent points results in a unique cycle, then G is a tree.* Since joining any pair of nonadjacent points gives a cycle, the points must be connected by a path. Thus G is connected.

Another Proof of Cayley's Formula

There are several other elegant proofs of Cayley's formula. Here we'll give a combinatorial proof that uses a induction, and a strengthening of the induction hypothesis.

A collection of trees is, naturally, called a *forest*. Let $T_{n,k}$ be the number of labeled forests on $\{1,\dots,n\}$ consisting of k trees. Then $T_{n,1}$ is the number of labeled trees; we want to show that $T_{n,1} = n^{n-2}$.

Suppose we have a forest F with k trees; we can assume that the vertices $\{1,2,\dots,k\}$ are in different trees. Suppose that vertex 1 is adjacent to i nodes. Then if we delete vertex 1, the i neighbors together with $2,\dots,k$ give one vertex each in the components of a forest with $k-1+i$ trees. We can reconstruct the original forest by first fixing i, then choosing the i neighbors of 1, and then the forest of size $k-1+i$ on $n-1$ nodes. Therefore,

$$T_{n,k} = \sum_{i=0}^{n-k} \binom{n-k}{i} T_{n-1,k-1+i} \tag{1}$$

with the initial conditions $T_{0,0} = 1$ and $T_{n,0} = 0$.

Proposition. $T_{n,k} = kn^{n-k-1}$. In particular, the number of labeled trees is $T_{n,1} = T_n = n^{n-2}$.

Proof. For a given n, suppose this holds for $k-1 \geq 0$. Then using (1) we have

$$\begin{aligned}
T_{n,k} &= \sum_{i=0}^{n-k} \binom{n-k}{i} T_{n-1,k-1+i} \\
&\overset{\text{I.H.}}{=} \sum_{i=0}^{n-k} \binom{n-k}{i} (k-1+i)(n-1)^{n-1-k-i} \\
&= \sum_{i=0}^{n-k} \binom{n-k}{i} (n-1-i)(n-1)^{i-1} \\
&= \sum_{i=0}^{n-k} \binom{n-k}{i} (n-1)^{i} - \sum_{i=1}^{n-k} \binom{n-k}{i} i(n-1)^{i-1} \\
&= n^{n-k} - (n-k) \sum_{i=1}^{n-k} \binom{n-1-k}{i-1} (n-1)^{i-1} \\
&= n^{n-k} - (n-k) \sum_{i=0}^{n-1-k} \binom{n-1-k}{i} (n-1)^{i} \\
&= n^{n-k} - (n-k)n^{n-1-k} \\
&= kn^{n-1-k}
\end{aligned}$$

See [1] for other elegant, and very different proofs of this result.

Euler's Formula. If G is a connected plane graph with n vertices, e edges and f faces, then

$$n - e + f = 2$$

Proof. Let $T \subset E$ be a subset of edges that forms a spanning tree for G. Let G^* denote the dual graph of G, with edge set E^*.

Consider the set of edges $T^* \subset E^*$ in the dual graph that correspond to edges in $E \backslash T$. Then T^* connects all of the faces, since T does not have a cycle. Also, T^* does not contain a cycle; if it did it would separate some vertices of G inside the cycle from the vertices outside the cycle, which is impossible since T is connected. Thus, T^* is a spanning tree of G^*.

Now, the number of vertices in a tree is one larger than the number of edges. Therefore, $n = e_T + 1$. Similarly, $f = e_{T^*} + 1$. Combining these gives

$$n + f = (e_T + 1) + (e_{T^*} + 1) = e + 2.$$

Corollary. Suppose that G is a plane graph with $n > 2$ vertices. Then

(a) G has a vertex of degree at most 5;

(b) G has at most $3n - 6$ edges.

Proof. Let f_i be the number of faces with i sides, and let n_j be the number of nodes with j neighbors. Then

$$\begin{aligned} f &= f_1 + f_2 + f_3 + f_4 + \cdots \\ n &= n_0 + n_1 + n_2 + n_3 + \cdots \end{aligned}$$

Moreover, since every edge has two endpoints, we have that

$$2e = n_1 + 2n_2 + 3n_3 + 4n_4 + \cdots$$

That is, every edge contributes 2 to the sum of all degrees. Similarly, we have that

$$2e = f_1 + 2f_2 + 3f_3 + 4f_4 + \cdots$$

That is, every edge borders two faces.

Now, since every face must have at least 3 sides, we have that in fact

$$\begin{aligned} f &= f_3 + f_4 + f_5 + \cdots \\ 2e &= 3f_3 + 4f_4 + 5f_5 + \cdots \end{aligned}$$

and therefore $2e - 3f \geq 0$.

To prove (a), suppose on the contrary that every vertex had degree at least 6. Then we would have

$$\begin{aligned} n &= n_6 + n_7 + n_8 + \cdots \\ 2e &= 6n_6 + 7n_7 + 8n_8 \end{aligned}$$

which would imply that $2e - 6n \geq 0$. Combining these gives

$$(2e - 6n) + 2(2e - 3f) = 6(e - f - n) \geq 0$$

which implies $e \geq n + f$, which contradicts Euler's formula.

To prove (b), we use the relation $2e - 3f \geq 0$ and Euler's formula to conclude that

$$3n - 6 = 3e - 3f \geq e.$$

Corollary. Every plane graph can be 6-colored.

Proof. By induction on the number of nodes. For small cases with fewer than 6 nodes, this is obvious. Suppose that the statement is true for planar graphs with fewer than n nodes, and let G be a plane graph with n nodes. Then G has a vertex v that has degree no larger than 5. Removing v from G, the resulting graph $G' = G - \{v\}$ is 6-colorable, by the induction hypothesis. But since v has no more than 5 neighbors, we can extend the coloring to all of G by coloring v different from its neighbors.

References

1. Martin Aigner and Günter M. Ziegler, *Proofs from THE BOOK*, Springer-Verlag, 2004.

2. Frank Harary, *Graph Theory*, Addison Wesley, 1995.