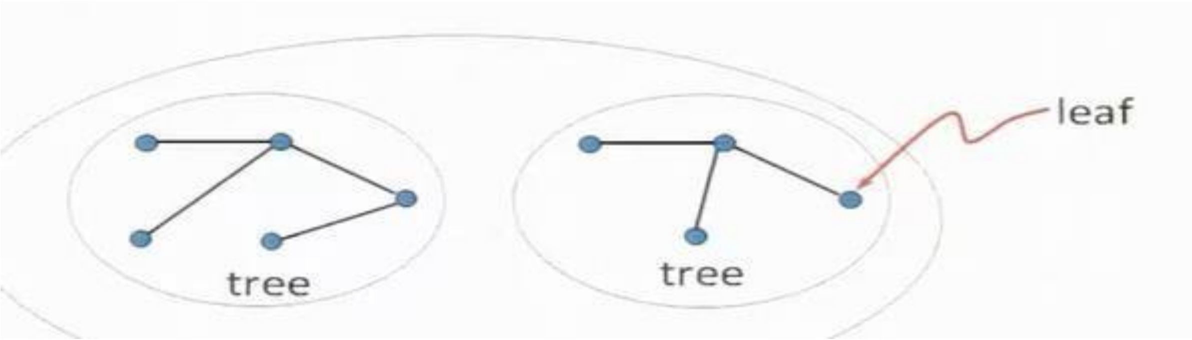
Trees

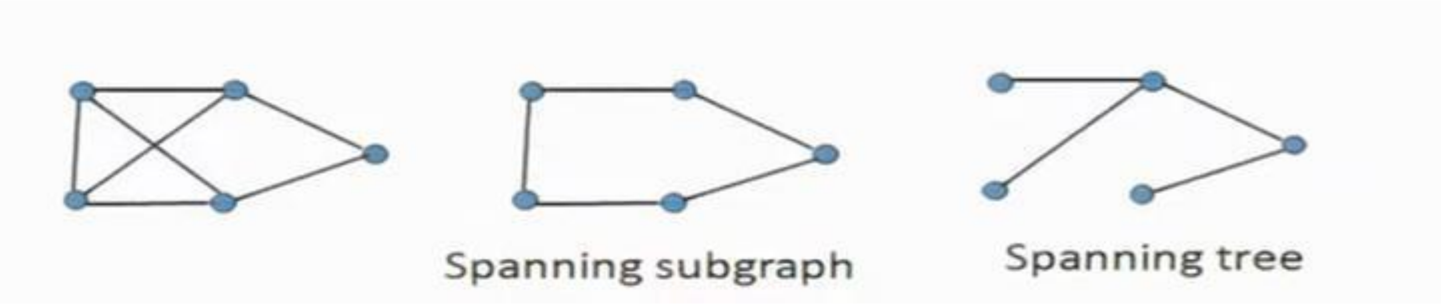
Definitions

- A graph with no cycle is **acyclic**.
- **A forest** is an acyclic graph.
- **A tree** is a connected acyclic graph.
- A **leaf** (or pendant vertex) is a vertex of degree **1**.

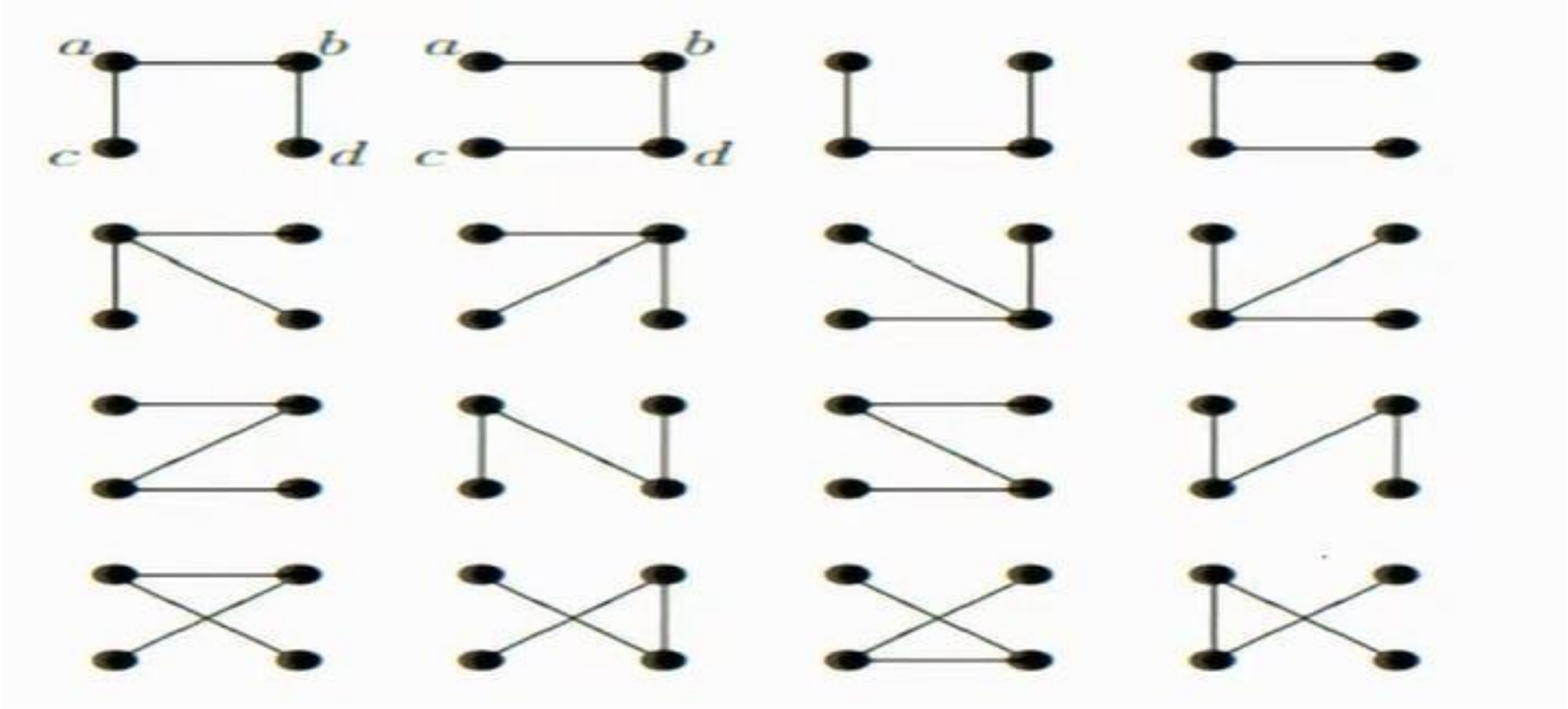


Spanning Subgraph

- A **spanning subgraph** of G is a subgraph with vertex set V(G)
- A **spanning tree** is a spanning subgraph that is a tree.



Spanning trees of K4

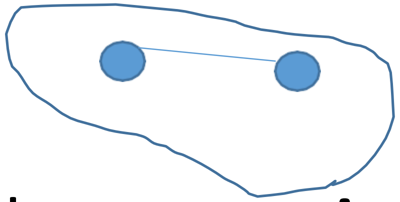


Important facts

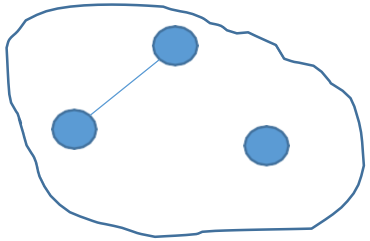
- **A tree is a connected forest**, and **every component** of a forest is a tree.
- A graph with no cycles has no odd cycles; **hence trees and forests are bipartite.**
- Paths are trees. A tree is a path if and only if its maximum degree is 2.
- **A star** is a tree consisting of one vertex adjacent to all the others. The n-vertex star is the bipartite $K_{1,n-1}$.
- **A graph that is a tree** has exactly **one spanning tree**: the full graph itself.

Important facts

- **A spanning subgraph of G need not be a connected**
- For example:
- If n(G)>1, then the empty subgraph with vertex set (V(G)) and edge set NULL is a spanning subgraph.



- **Connected subgraph of G** need not be a **spanning subgraph**.
- If n(G) > 2, then a subgraph consisting of one edge and its endpoints is connected. Now here is a connected forest this particular.



Lemma

Lemma: Every tree with at least two vertices has at least two leaves.

Proof:

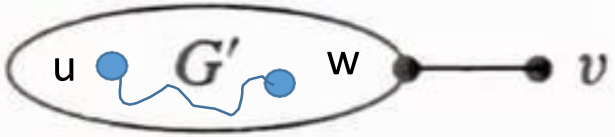
- A connected graph with **at least two vertices has an edge**.
- In an acyclic graph, **an endpoint of a maximal nontrivial path** has no neighbor other than its neighbor on the path.
- Hence the endpoints of such a path are leaves.

Cannot be Extended

Lemma: Deleting a leaf from a tree produces a tree with (n - 1) vertices.

Proof:

- Let v be a leaf of a tree G, and let $G' = G - v$.
- **A vertex of degree 1 belongs to no path connecting two other vertices.**
- **Therefore, for $u, w \in V(G')$, every u, w-path in G is also in G'.**
- Hence G' is connected.



- Since deleting a vertex cannot create a cycle, G' also is acyclic.
- Thus G' is a tree with n-1 vertices.

This Lemma implies that every tree with more than one vertex arises from a smaller tree by adding a vertex of degree 1

Lemma: A tree of n nodes has (n - 1) links

Base case:

- Tree with $n = 1$
- $m = 0,\ e = 0$ (a tree with one vertex has no edges)

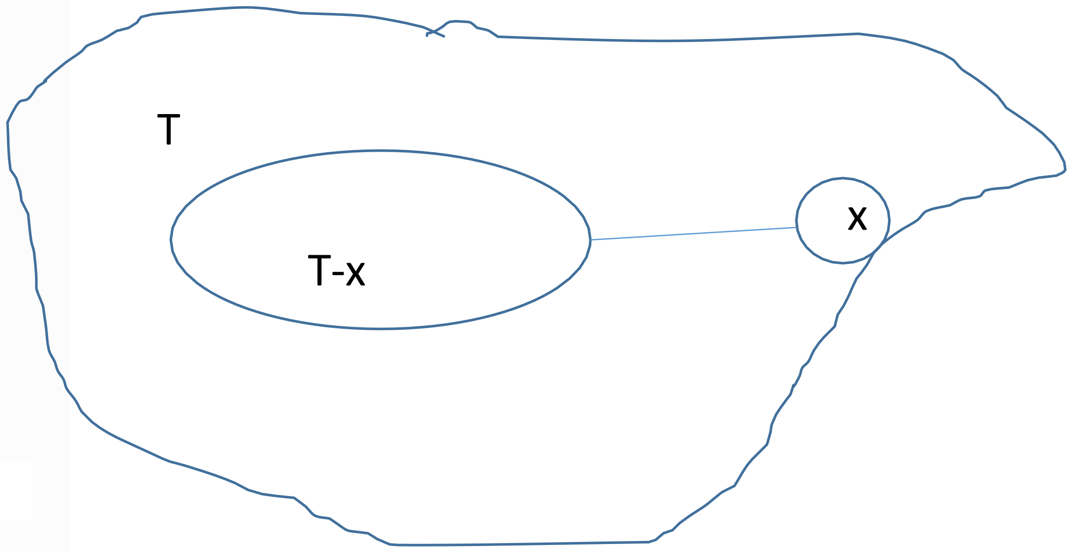
Induction hypothesis:

Suppose every tree with k vertices has $k - 1$ edges, where $k \geq 1$.

Induction step:

- Let T be a tree with $k + 1$ vertices.
- Since $k + 1 \geq 2$, T has a leaf x.
- The graph $T - x$ is a tree with k vertices
- By the induction hypothesis, $T - x$ has exactly $k - 1$ edges.
- Since T has one more edge than $T - x$, the tree T has exactly k edges.
- Hence, a tree with $k + 1$ vertices has $(k + 1) - 1 = k$ edges.

Therefore, **every tree with n vertices has exactly $n - 1$ edges.**



Theorem

Theorem: For an n-vertex graph G (with (n >= 1)), the following statements are equivalent (and characterize the trees with (n) vertices):

A) G is connected and has no cycles.

B) G is connected and has (n - 1) edges.

C) G has (n - 1) edges and no cycles.

D) G has no loops and has, for each (u, v \in V(G)), exactly one (u,v)-path.

Proof: We first demonstrate the equivalence of A, B, and C by proving that
any two of [connected, acyclic, (n - 1) edges] together imply the third.

Theorem

A: (G) is connected and has no cycles.

Connected, acyclic → (n - 1) edges.

Proof

We use **induction on number of nodes (n).**

- For (n = 1), an acyclic 1-vertex graph has no edge.
- **Induction hypothesis:** For (n > 1), suppose the implication holds for graphs with **fewer than n vertices.**
- Given an acyclic connected graph (G), we detect a leaf (v) and state that (G' = G - v) is also acyclic and connected.
- Applying the induction hypothesis to (G') yields

  e(G') = n - 2.

n-1

Since only one edge is incident to (v), we have that

e(G) = e(G') + 1 = n - 1.

Theorem

Theorem: For an n-vertex graph G (with (n >= 1)), the following statements are equivalent (and characterize the trees with (n) vertices):

A) G is connected and has no cycles.

B) G is connected and has (n - 1) edges.

C) G has (n - 1) edges and no cycles.

D) G has no loops and has, for each (u, v \in V(G)), exactly one (u,v)-path.

Proof: We first demonstrate the equivalence of A, B, and C by proving that
any two of [connected, acyclic, (n - 1) edges] together imply the third.

Theorem

B:G is connected and has (n - 1) edges.

- Connected and (n - 1) edges → acyclic.

Proof: Assume G (of n nodes) has cycle

- If (G) has cycle, **delete edges from cycles of (G)** one by one until the resulting **graph (G') is acyclic.**
- Since **no edge of a cycle** is a **cut-edge → (G') is connected**.
- From previous results **(A)**, it follows that
- e(G') = n - 1.
- Since we are given e(G) = n - 1, no edges were deleted.
- Thus (G' = G), and (G) is acyclic.

Theorem

Theorem: For an n-vertex graph G (with (n >= 1)), the following statements are equivalent (and characterize the trees with (n) vertices):

A) G is connected and has no cycles.

B) G is connected and has (n - 1) edges.

C) G has (n - 1) edges and no cycles.

D) G has no loops and has, for each (u, v \in V(G)), exactly one (u,v)-path.

Proof: We first demonstrate the equivalence of A, B, and C by proving that
any two of [connected, acyclic, (n - 1) edges] together imply the third.

Theorem

C: **(G) has (n - 1) edges and no cycles.**

(n - 1) edges and no cycles → connected.

Proof

- Let $G_1, \ldots, G_k$ be the components of G.
- Since every vertex appears in one component,

$$\sum_{i=1}^{k} n(G_i) = n.$$

- Since G has no cycles, each component satisfies the previous result; hence,

$$e(G_i) = n(G_i) - 1.$$

- Summing over i yields

$$e(G) = \sum_{i=1}^{k} e(G_i) = \sum_{i=1}^{k} \big(n(G_i) - 1\big) = n - k.$$

- Since $e(G) = n - 1$, it follows that $k = 1$.
- Therefore, G is **connected**.

Theorem

Theorem: For an n-vertex graph G (with $(n \geq 1)$), the following statements are equivalent (and characterize the trees with (n) vertices):

A) G is connected and has no cycles.

B) G is connected and has $(n - 1)$ edges.

C) G has $(n - 1)$ edges and no cycles.

D) G has no loops and has, for each $(u, v \in V(G))$, exactly one (u,v)-path.

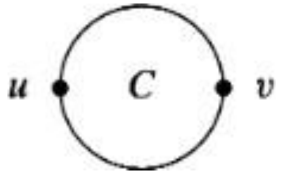
Proof: We first demonstrate the equivalence of A, B, and C by proving that
any two of [connected, acyclic, $(n - 1)$ edges] together imply the third.

D ⇒ A

For every (u, v \in V(G)), one and only one (u,v)-path exists → (G) is connected and acyclic.

Proof

- If there is a (u,v)-path for every (u,v \in V(G)), then (G) is connected.
- If (G) has a **cycle (C)**, then (G) has **two (u,v)-paths for (u, v \in V(C))**.
- Hence (G) is acyclic



Corollary

a) Every edge of a tree is a cut-edge.

Proof: A tree has **no cycles**, so every edge is a cut-edge.

b) Adding one edge to a tree forms exactly one cycle.

Proof: There is a **unique path linking each pair of vertices**, so joining two vertices by an edge creates **exactly one cycle.**

c) Every connected graph contains a spanning tree.

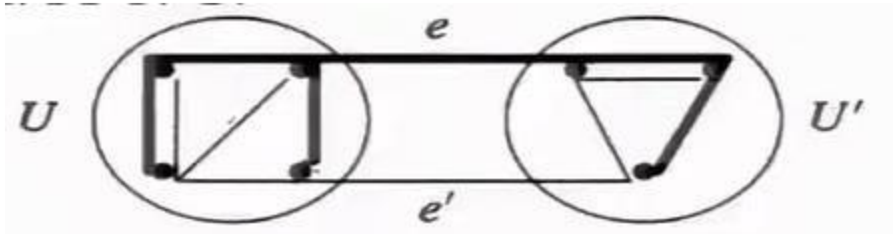
Proof: **Iteratively deleting edges from cycles in a connected graph yields** a connected spanning acyclic graph. Hence, the resulting graph is a tree.

Theorem

Theorem: If (T, T') are spanning trees of a connected graph (G), and e € E(T) - E(T')), then there is an edge e' € E(T') - E(T)) such that (T - e + e') is a spanning tree of G.

Proof

- **Every edge of (T) is a cut-edge of (T).**
- Let (U) and (U') be the two components of (T - e).
- Since (T') is connected, (T') has an edge (e') with endpoints in (U) and (U').
- Now (T - e + e') is connected, has (n(G) - 1) edges, and is a spanning tree of (G).



Theorem

Lemma: If T, T' are spanning trees of a connected graph G and e € E(T) - E(T'), then there is an edge e' € E(T') - E(T) such that T' + e - e' is a spanning tree of G

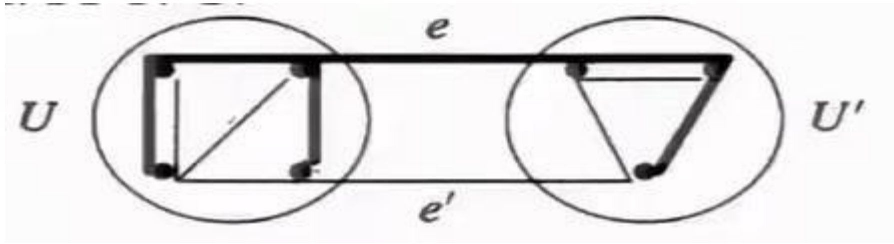
Proof:

The graph **T' + e contains a unique cycle C**.

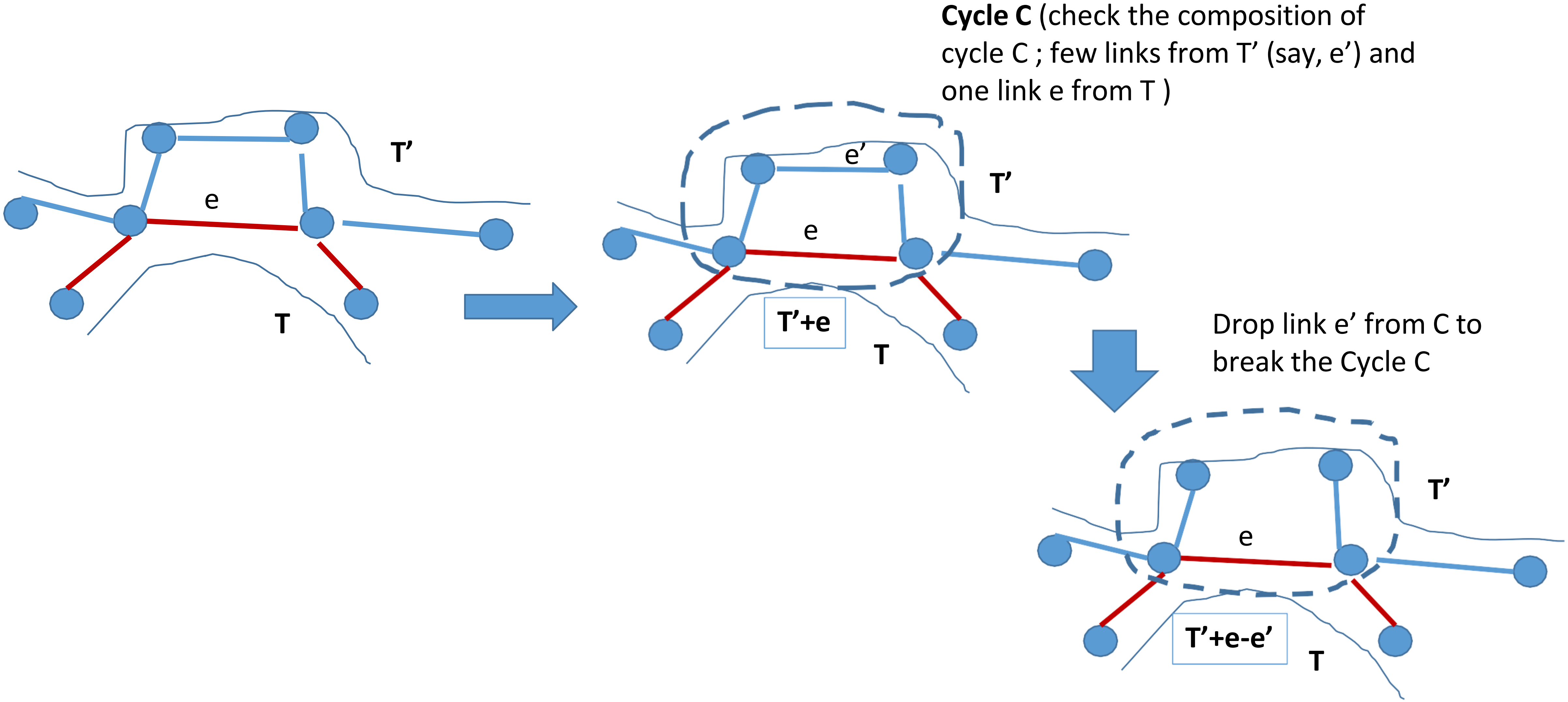
Since T is acyclic, there is an edge e' € E(C) - E(T).

Deleting e' breaks the **only cycle in T' + e**.

Now T' + e - e' is connected and acyclic and is a spanning tree of G.

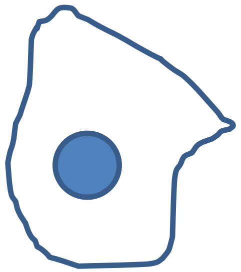


Lemma: If T, T' are spanning trees of a connected graph G and e € E(T) - E(T'), then there is an edge e' € E(T') - E(T) such that T' + e - e' is a spanning tree of G



Lemma

Lemma: If T is a tree with k edges and G is a simple graph with $\delta(G) >= k$, then T is a subgraph of G.

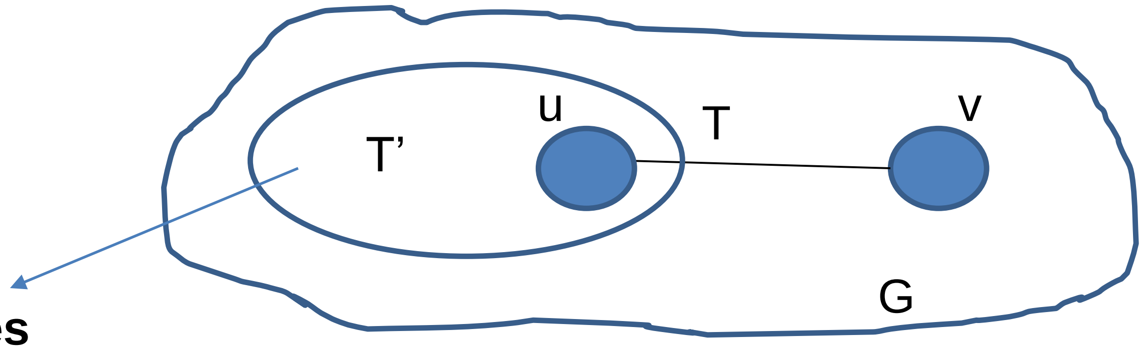


Proof: We use **induction** on **number of links k**

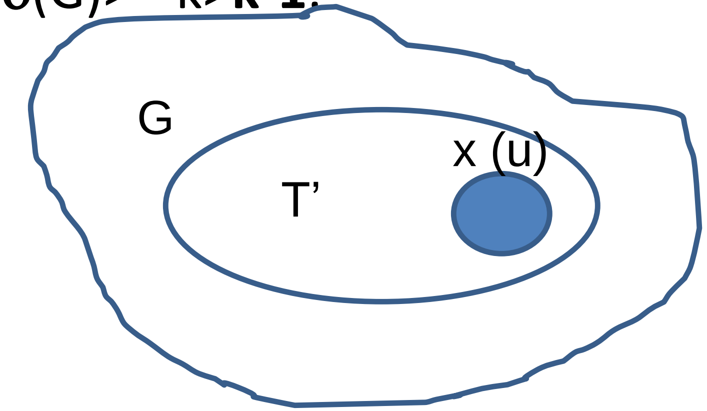
Basis step: (k=0). Every simple graph contains (K_1), which is the only tree with no edges.

Induction hypothesis: For k>0. We assume that the **claim holds for trees** with **fewer than (k) edges**.

- **Hence, a simple graph $\delta(G) >= k-1$ contains a tree with k-1 links**
- Consider a **tree T with k links**, choose a **leaf v** in T;
- Let **u be its neighbor**. Consider the **smaller tree T'=T-v**.
- By the **induction hypothesis**, graph **G contains T' as a subgraph**, since $\delta(G) >= k >$ **k-1**.



k-1 links, k nodes

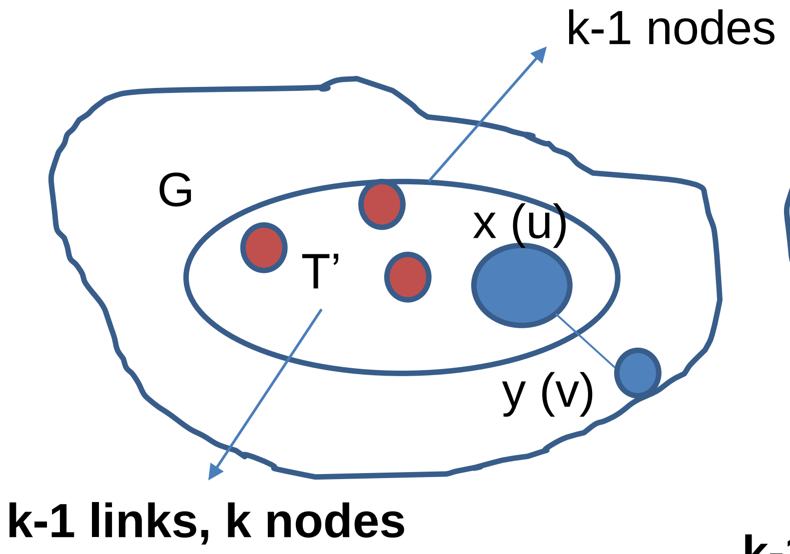


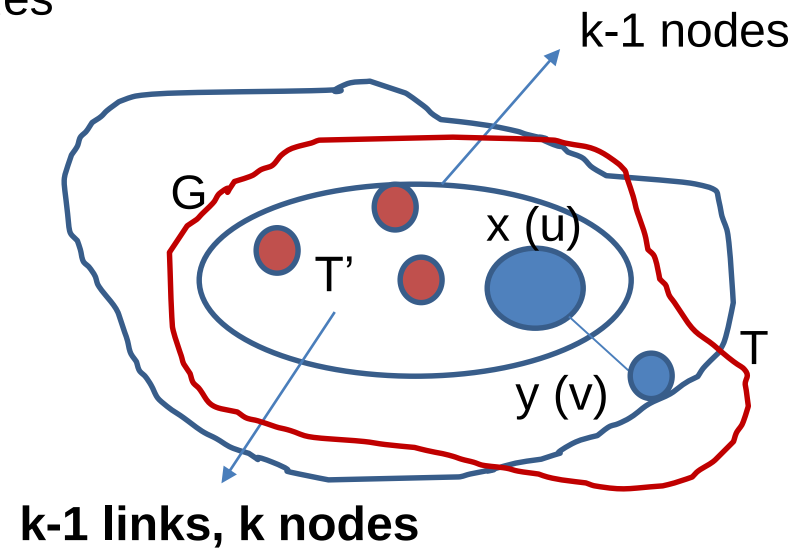
Lemma

Lemma: If T is a tree with k edges and G is a simple graph with δ(G)>= k, then T is a subgraph of G.

- Let **x be the vertex** in this copy of **T'** that **corresponds to u**.
- Because **T' has only (k-1) vertices** other than (u) and **deg(x)>= k**,
- **Vertex x has a neighbor y** in (G) that is not in this copy of (T') ; **one link of x-y goes outside T'**.

- **Adding the edge x-y expands this copy of (T') into a copy of (T)**, with (y) corresponding to (v).
- v is outside T', but part of T. **Hence T is inside graph G**





Diameter, Eccentricity, Radius

- If G has a u, v-path, then the **distance** from u to v, written $d_G(u, v)$ or simply $d(u, v)$, is the least length of a u, v-path. If G has no such path, then $d(u, v) = \infty$.
- The **diameter** ($\operatorname{diam} G$) is

$$\max_{u,v \in V(G)} d(u, v).$$

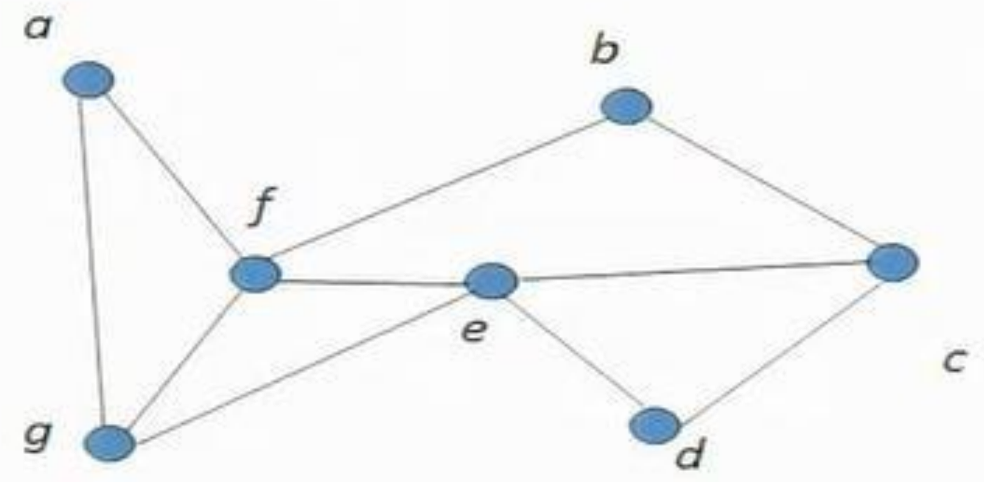
  - Upper bound of the distance between every pair.
- The **eccentricity** of a vertex u, written $\varepsilon(u)$, is

$$\max_{v \in V(G)} d(u, v).$$

  - Upper bound of the distance from u to the others.
- The **radius** of a graph G, written $\operatorname{rad} G$, is

$$\min_{u \in V(G)} \varepsilon(u).$$

  - Lower bound of the eccentricity.



Distance(f,c) : 2
Distance(g,c): 2
Distance(a,c): 3

diameter: 3

eccentricity(f):2
eccentricity(a): 3

radius: 2

Diameter, Eccentricity, Radius

- The **diameter** equals the **maximum** of the vertex **eccentricities**

- In a **disconnected graph**, the diameter and radius (and every eccentricity) are **infinite**, because the distance between vertices in different components is infinite.

Diameter, Eccentricity, Radius

- Cycle C_n has diameter ceil(n/2)

Each vertex has the **same eccentricity**

Diam(C_n)=Raid(C_n)

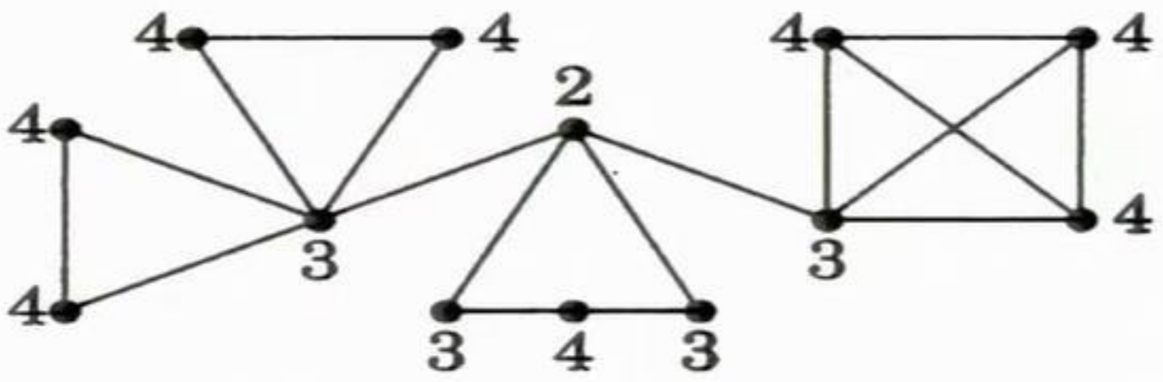
- For (n >= 3), the **n-vertex tree of least diameter is the star**, with diameter 2 and radius 1.

- The **largest diameter is the path**, with diameter (n-1)
- Largest radius is ceil((n-1)/2).

- **Every path in a tree** is the shortest (the only) path between its endpoints.
- So the diameter of a tree is the length of its longest path.

Examples

- In the graph below, each vertex is labeled with its eccentricity. The radius is 2, the diameter is 4, and the length of the longest path is 7.



Theorem

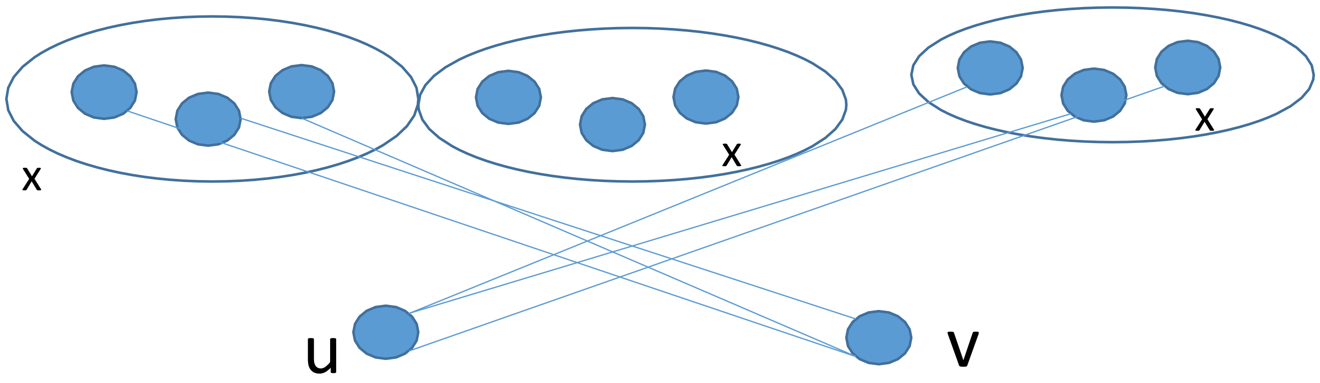
Theorem:

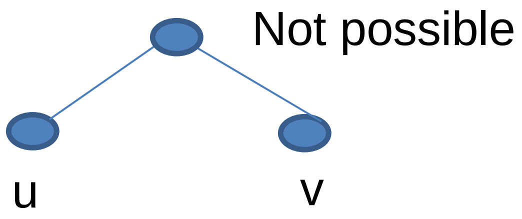
If (G) is a simple graph, then Diam(G)>= 3 → Diam(G')<= 3.

Proof:

When Diam(G)>=3, there exist **nonadjacent vertices (u,v ∈ V(G)) with no common neighbour**

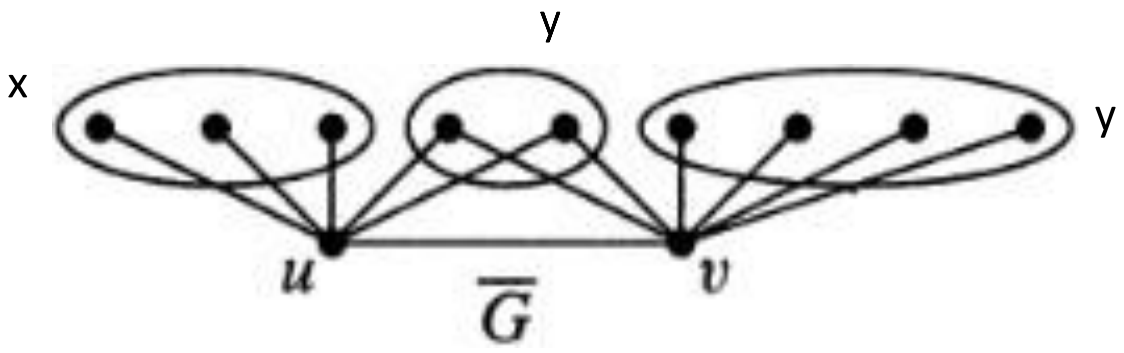
Hence every **x** ∈ V(G)-{u,v}) has **at least one of {u,v} as a nonneighbor**.





Theorem

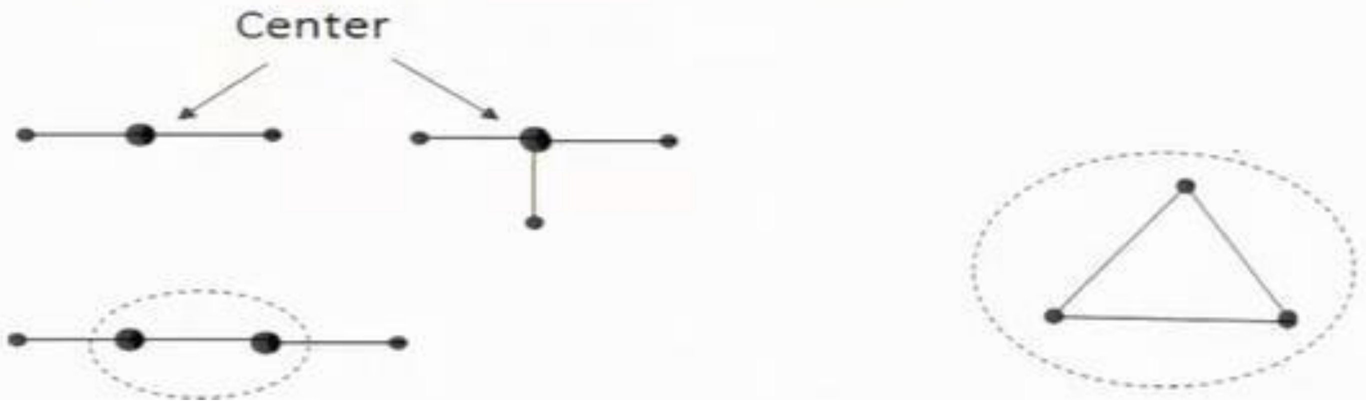
- **In G', (u, v) becomes adjacent**
- Moreover, **x becomes adjacent to at least one of {u,v} in G'**.
- Since also (uv € E(G'), for every pair (x,y) there is an (x,y)-path of length at most 3 in graph G' through ({u,v}).
- Hence, diam(G') <= 3.



Center of a Graph

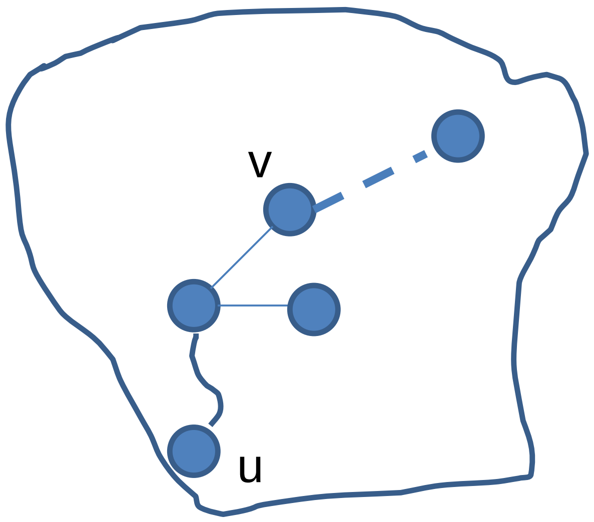
Definition: The **center** of a graph G is the subgraph induced by the vertices of minimum eccentricity.

- The center of a graph is the full graph if and only if the radius and diameter are equal.



If u be a vertex in a tree T, every vertex farthest from u is a leaf.

Contradiction



Theorem

Theorem : The center of a tree is a vertex or an edge.

Proof:

- We use **induction on the number of vertices** in a tree (T).
- **Basis step**: (n(T) <= 2). With at most two vertices, the center is the entire tree.

Theorem

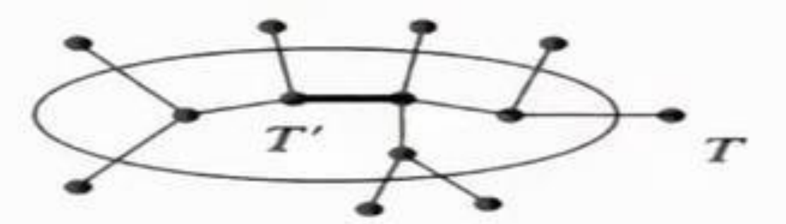
Induction step: $n(T) > 2$

- Form T' by deleting every leaf of T. **Hence,** T' is a tree.
- Since the internal vertices on the paths between leaves of T remain, T' has at least one vertex.
- Every vertex at maximum distance in T from a vertex $u \in V(T)$ is a leaf (otherwise, the path reaching it from u can be extended farther).
- Since all the leaves have been removed and no path between two other vertices uses a leaf,

$$\varepsilon_{T'}(u) = \varepsilon_T(u) - 1 \quad \text{for every } u \in V(T').$$

- Also, the eccentricity of a leaf in T is greater than the eccentricity of its neighbor in T.
- Hence the vertices minimizing $\varepsilon_T(u)$ are the same as the vertices minimizing $\varepsilon_{T'}(u)$.

Could a deleted leaf be a center?



Spanning trees

- There are $2^{\binom{n}{2}}$ simple graphs with vertex set $[n]=\{1,\ldots,n\}$, since each pair may or may not form an edge.
- How many of these are trees?
- With vertex set $[n]$, there are n^{n-2} labelled trees; this is **Cayley's Formula**.

Spanning trees

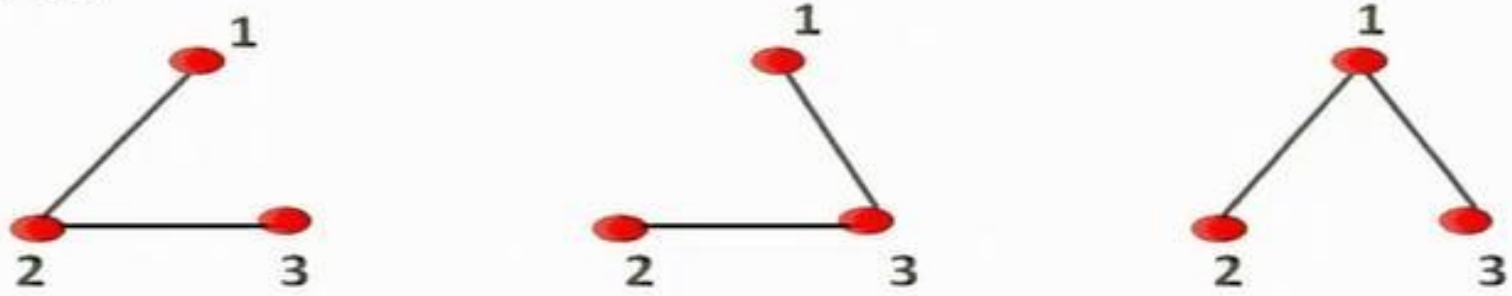
- **Cayley's Formula** tells us how many different trees we can construct on *n* vertices. These are called spanning trees on n vertices. There are n^{n-2} trees on a vertex set V of n elements.

- In its simplest form, Cayley's Formula says:
- $|T_n| = n^{n-2}$

- **Example:**
- $|T_2| = 2^{2-2} = 1,\ |T_3| = 3^{3-2} = 3,\ |T_4| = 4^{4-2} = 16$

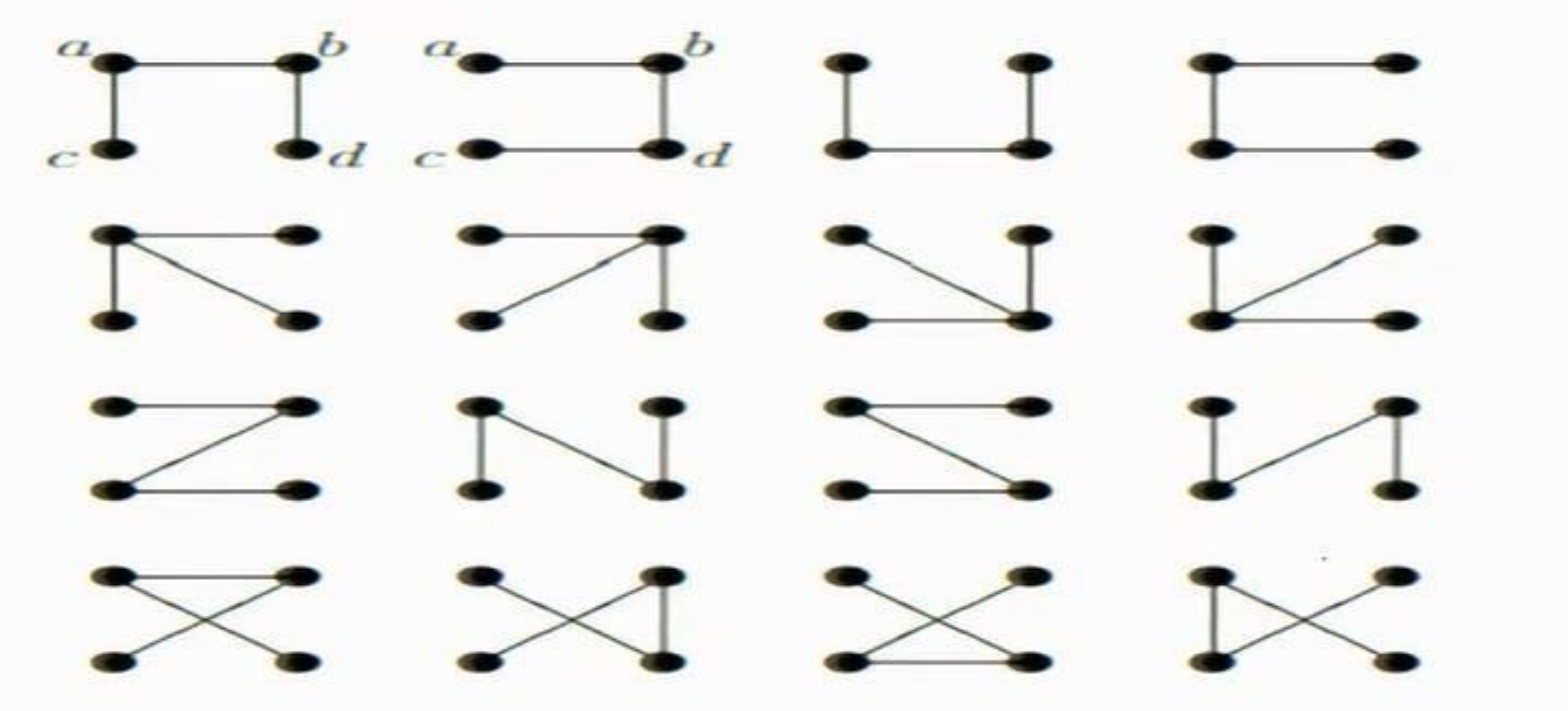
Cayley's formula

Example: $n = 3$



- With one or two vertices, only one tree can be formed.
- With three vertices there is till only one isomorphism class, but there are three trees with vertex set [3].
- With vertex set [4], there are four stars and 12 paths, yielding 16 trees.
- With vertex set [5], a careful study yields 125 trees.

Spanning trees of T4



Cayley's formula

- Now we may see a pattern. With vertex set *[n]*, there are n^{n-2} trees; this is **Cayley's Formula**. Prüfer, Kirchhoff, Pólya, Renyi, and others found proofs.
- We will discuss a bijective proof, establishing a one-to-one correspondence between the **set of trees with vertex set** ***[n]*** and **a set of known size.**
- Given a set S on n numbers, there are exactly n^{n-2} ways to form a list of length $n-2$ with entries in S. The set of lists is denoted S^{n-2} .
- We use S^{n-2} to encode the trees with vertex set S. The list that results from a tree is its **Prüfer code.**

Proof Idea: Labelled tree $T \leftrightarrow$ Prüfer sequence S

1:1 correspondence

on n elements of length n-2

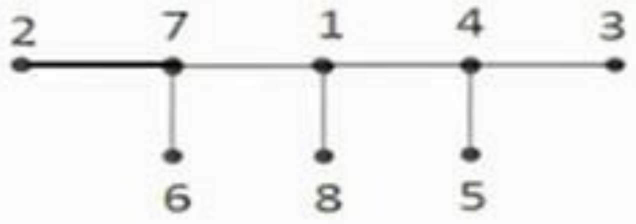
- How many such sequence ? $n \times n \times n \times \ldots\ldots \times n = n^{n-2}$

Prufer code

Algorithm. (Prüfer code) Production of $f(T) = (a_1,\ldots,a_{n-2})$

Input: A tree T with vertex set $S \subseteq N$

Iteration: At the ith step, delete the least remaining leaf, and let a_i be the ***neighbor*** of this leaf.



1. Delete 2 $a_1 = 7$
2. Delete 3 $a_2 = 4$
3. Delete 5 $a_3 = 4$
4. Delete 4 $a_4 = 1$
5. Delete 6 $a_5 = 7$
6. Delete 7 $a_6 = 1$

Prufer code

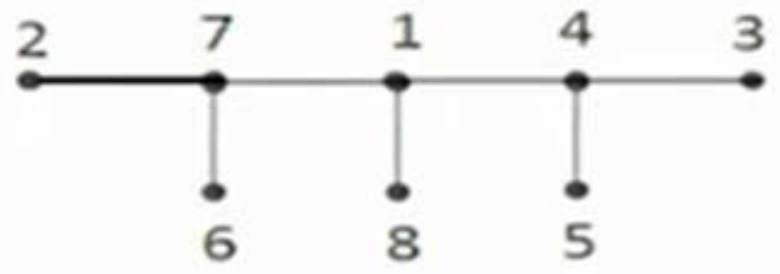
Algorithm: Labelled tree T → Prüfer Sequence S

- Find a leaf of T with smallest label

  → add the neighbour to sequence S

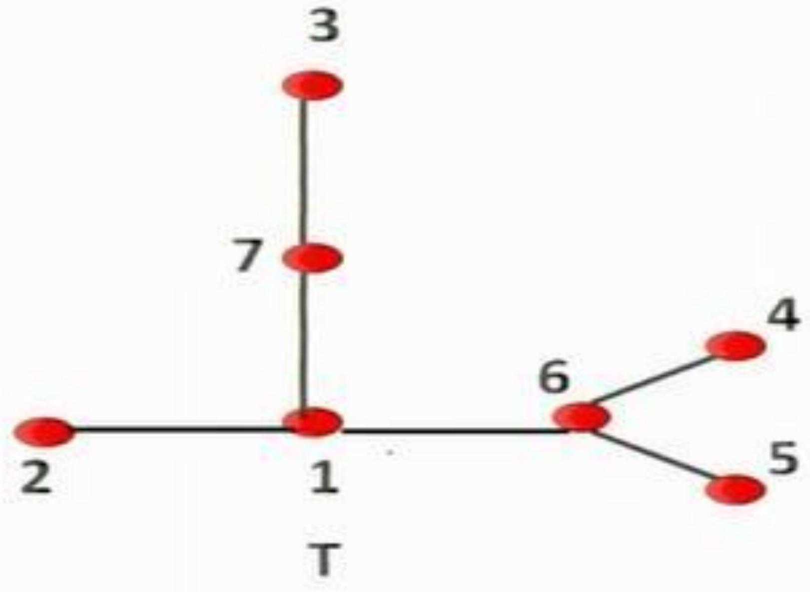
  → delete this leaf

- Repeat until our tree is K_2



1. Delete 2 $a_1 = 7$
2. Delete 3 $a_2 = 4$
3. Delete 5 $a_3 = 4$
4. Delete 4 $a_4 = 1$
5. Delete 6 $a_5 = 7$
6. Delete 7 $a_6 = 1$

Prufer code



S = ( )

S = (1)

S = (1, 7)

S = (1, 7, 6)

S = (1, 7, 6, 6)

S = (1, 7, 6, 6, 1)

Why Prufer sequence length (n-2)?

- No leaf gets appended to *S*
- Every vertex *v* is added to *S* a total of *deg(v)-1* times

So in given Tree let n vertices, *m* edges $\Rightarrow$ $m = n - 1$

Number of terms in *S*

$$\sum_{v \in v(T)} (\deg(v) - 1) = \left(\sum_{v \in v(T)} \deg(v) \right) - \sum_{v \in v(T)} (1)$$

$$= 2m - n$$

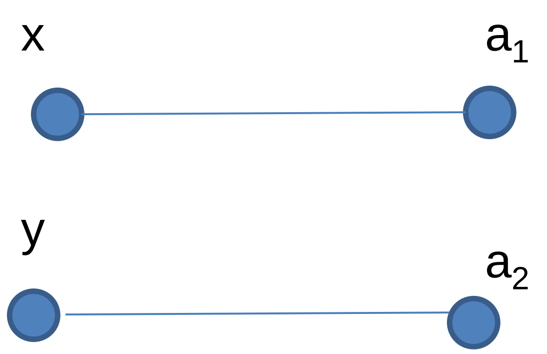
$$= 2\,(n\text{-}1) - n$$

$$= n - 2$$

Generation of Labelled Tree from Prüfer Sequence

Prüfer sequence (S) of length (n-2) using symbols (1,..,n) → Labelled tree (T)

Let $S=(a_1,a_2,\ldots\ldots,a_{n-2})$, $a_i \in \{1,\ldots,n\}$
Set of nodes in T, $L=\{1,\ldots,n\}$
Algorithm: Prüfer Sequence (S) → Labelled Tree (T)



Find smallest element ($x \in L=\{1,\ldots,n\}$), ($x \notin S$)

- Join x to 1st element of (S)
- Delete a_1 from S, delete x from $L=\{1,\ldots,n\}$

$L' = L - \{x\}$, $S' = S - \{a_1\}$

Find smallest ($y \in L'$) not in S'
Join y to the 1st element of S'
Delete a_2 from S'
Delete (y) from L'
Continue until 2 items remain in L'

Example

S = (1, 7, 6, 6, 1)   length 5 → n = 7

L = {1, 2, ..., 7}- vertices with labels

S → T

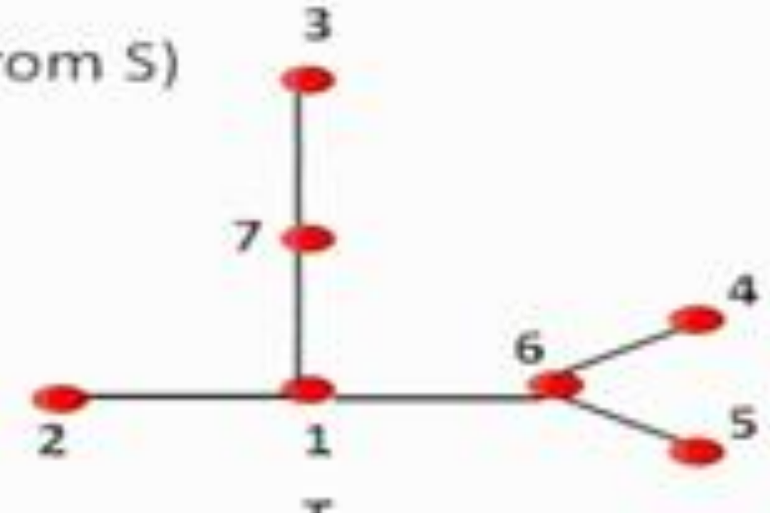
T → S

x ∈ *L \ {S} smallest* → x = 2

(edge 12, remove 2 from L and 1 from S)

S = (7, 6, 6, 1)

L = (1, 3, 4, 5, 6, 7)



Cayley's Formula

Theorem: For a set ($S \subseteq N$) of size (n), there are (n^{n-2}) trees with vertex set S.

Proof: We assume (n>= 2). We prove that it defines a bijection function (f) **from the set of trees with vertex set (S)** to the **set (S^{n-2}) of lists of length (n-2) from S**.

We must show for each a=(a_1,…..,a_{n-2}) $\in S^{n-2}$
that **exactly one tree (T)** with vertex set (S) satisfies f(T)=a

We can prove this by induction on (n).

Proof by Induction:

Basis step: (n=2): There is one tree with two vertices. The Prüfer code is a list of length (0), and it is the only such list. **(Try for n=3)**

Induction step: (n>2):

Computing f(T) reduces each vertex to degree 1 and then possibly deletes it.

Thus every nonleaf vertex in (T) appears in f(T).

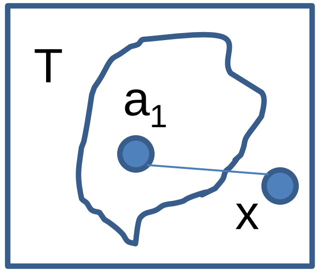
No leaf appears, because recording a leaf as a neighbor of a leaf would require reducing the tree to one vertex.

Hence the **leaves of T** that are the elements of S **not in f(T)**.

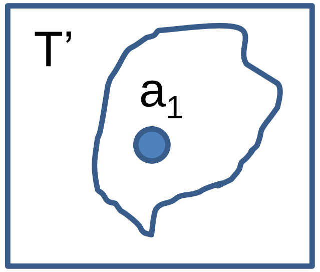
If f(T)=a, **a = $(a_1,\ldots\ldots,a_{n-2}\})$**

then the first leaf deleted is the **smallest element of S** not in **a** (call it x), and the neighbor of x is a_1.

x smallest
x not in **a**



$a= (a_1,\ldots\ldots,a_{n-2})$



$a'= (a_2,\ldots\ldots,a_{n-2})$

S'=S-{x}

- We have shown that every such tree has **x** as its **least leaf** and has the edge **(x-a_1)**.
- Deleting (x) leaves a tree with vertex set (S'=S-{x}).
- Its Prüfer code is a'=(a_2,…….,a_{n-2}), an (n-3)-tuple formed from (S').

- By the induction hypothesis, there exists exactly one tree (T') having vertex set (S') and Prüfer code (a').

- Since every tree with Prüfer code (a) is formed by adding the edge (x-a_1) to such a tree, there is exactly **one** tree to (f(T)=a).

- Furthermore, adding the edge (x-a_1) to (T') does produce a tree with vertex set (S) and Prüfer code (a), so there is exactly **one** solution.

Therefore, there is exactly one tree (T) such that
f(T)=a.

Hence (f) is a bijection, which proves Cayley's formula: \boxed{n^{,n-2}} trees on a given set of (n) labelled vertices.

Count the trees by their vertex degrees

Given positive integers d_1, $d_{2,..}$, d_n summing to (2n – 2), there are exactly

$$\frac{(n-2)!}{\prod (di-1)!}$$

trees with vertex set [n] such that vertex i has degree d_i, for each i.

By the Handshaking Lemma,

$$\sum_{i=1}^{n} \deg(i) = 2|E|.$$

Therefore,

$$\sum_{i=1}^{n} d_i = 2(n-1) = 2n-2.$$

Count the trees by their vertex degrees

Proof:

- While **constructing** the Prüfer code **a** of a tree **T**,
- we **record x** each time we **delete a neighbor (leaf) of x**, until we delete x itself or leave x among the last two vertices.

- **Thus each vertex x appears $d_T(x)$ - 1 times in the Prüfer code a.**
- **Hence the elements present in Prufer code a is fixed for all the trees T with degree $d_1, \ldots, d_n$**

- We **count trees with these vertex degrees** by **counting lists of length n - 2**, where each vertex i appears $(d_i - 1)$ times in **a**

- We are **permuting n - 2 distinct objects** → there are **$(n - 2)!$ lists** (If we assign subscripts to the copies of each i to distinguish them)

- Since the **copies of i are not distinguishable**, we have counted each desired arrangement $(d_i - 1)!$ times, once for each way to order the subscripts on each type of label.

Example

- ***Trees with fixed degrees.*** Consider trees with vertices $\{1,2,3,4,5,6,7\}$ that have degrees $(3,1,2,1,3,1,1)$, respectively. We compute $\frac{(n-2)!}{\prod (di-1)!} = 30$; The trees are suggested below
- Only the vertices $\{1,3,5\}$ are non-leaves. Deleting the leaves yields a subtree on $\{1,3,5\}$. There are three such subtrees, determined by which of the three is in the middle.