Lecture 05
Trees and Distance

Dr. Rajiv Misra
Associate Professor
Dept. of Computer Science & Engg.
Indian Institute of Technology Patna
rajivm@iitp.ac.in

Preface

Recap of previous Lecture:

- In the previous lecture, we have discussed the application of eulerian circuit i.e. chinese postman problem, degree sequence, graphic sequences, 2-switch and theorems based on graphic sequences.

Content of this Lecture:

- In this lecture, we will discuss the basic properties of trees and use of distance in trees and graphs.

Basic Properties

- The word "**tree**" suggests branching out from a root and never completing a cycle.

- Trees as graphs have many applications, especially in:

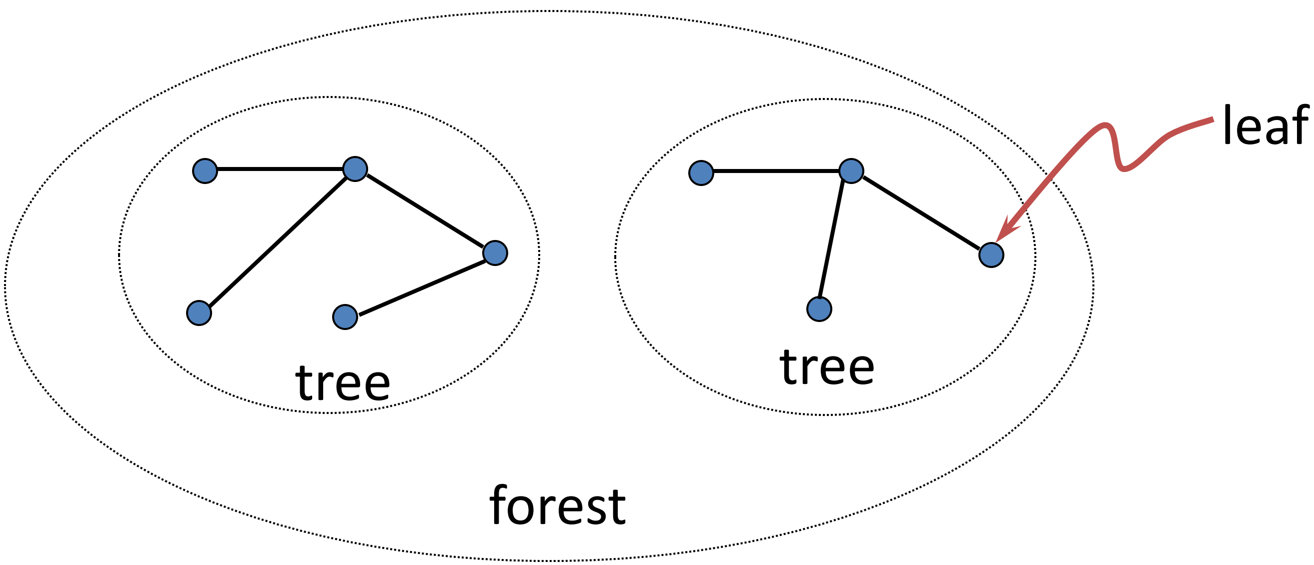
(i) Data storage

(ii) Searching and

(iii) Communication

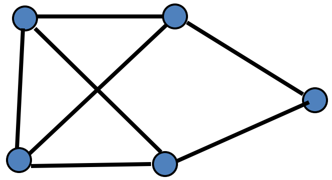
Definitions

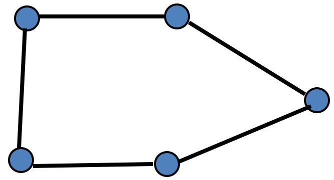
- A graph with no cycle is ***acyclic***
- A ***forest*** is an acyclic graph
- A ***tree*** is a **connected** acyclic graph
- A ***leaf*** (or ***pendant vertex***) is a vertex of degree 1



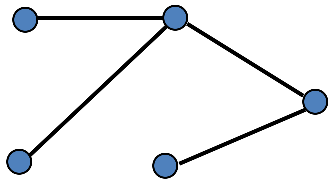
Spanning Subgraph 2.1.1

- 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 subgraph



Spanning tree

Example

- 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 biclique $K_{1,n-1}$
- A graph that is a tree has exactly one spanning tree; the full graph itself. A spanning subgraph of G need not be connected, and a connected subgraph of G need not be a spanning subgraph. For example:
- **If n(G) > 1**, then the empty subgraph with vertex set V(G) and edge set Ø is spanning but not connected.
- **If n(G) > 2**, then a subgraph consisting of one edge and its endpoints is connected but not spanning. ■

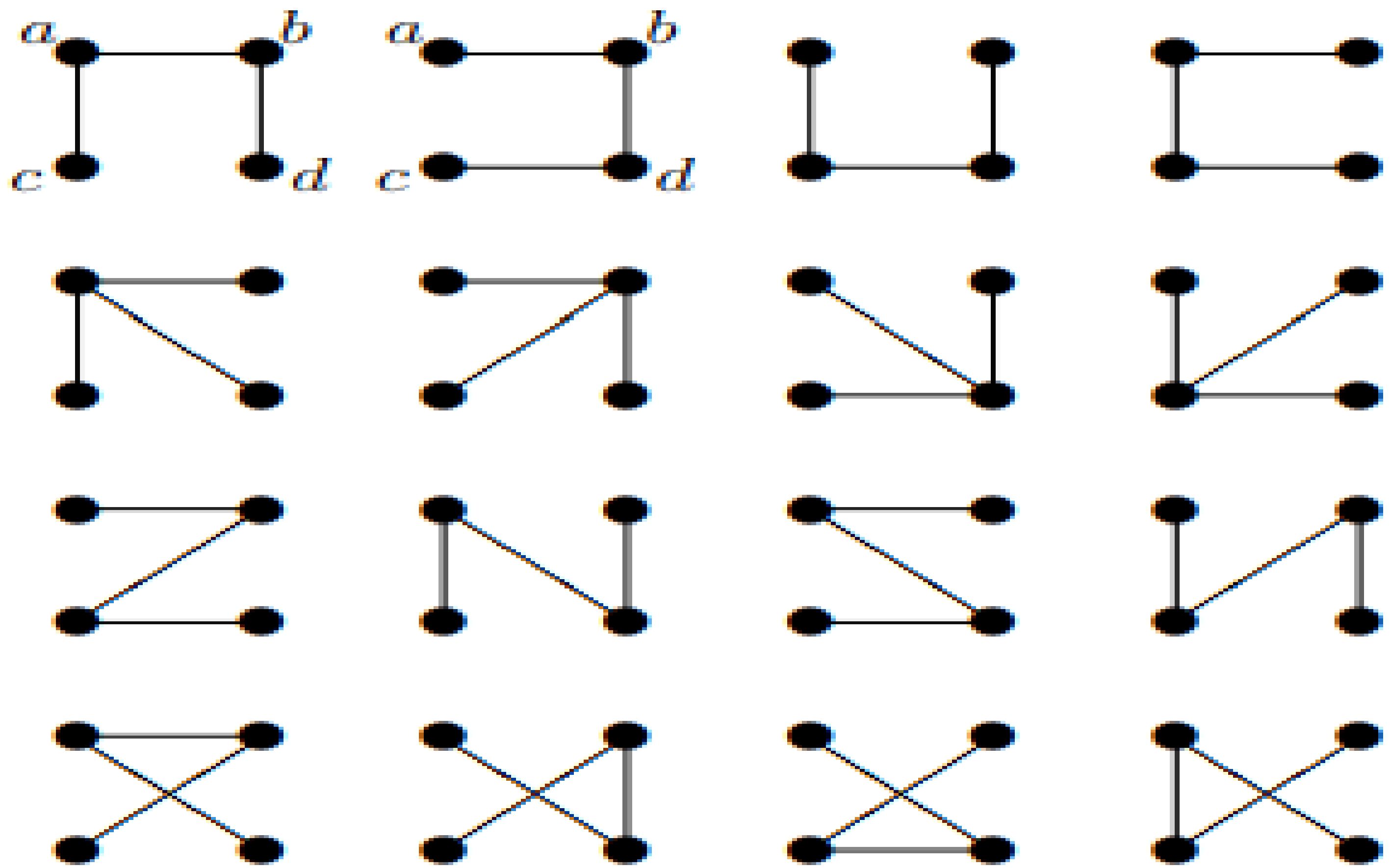
Cayley's Formula

- **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:
- $|T_n| = n^{n-2}$

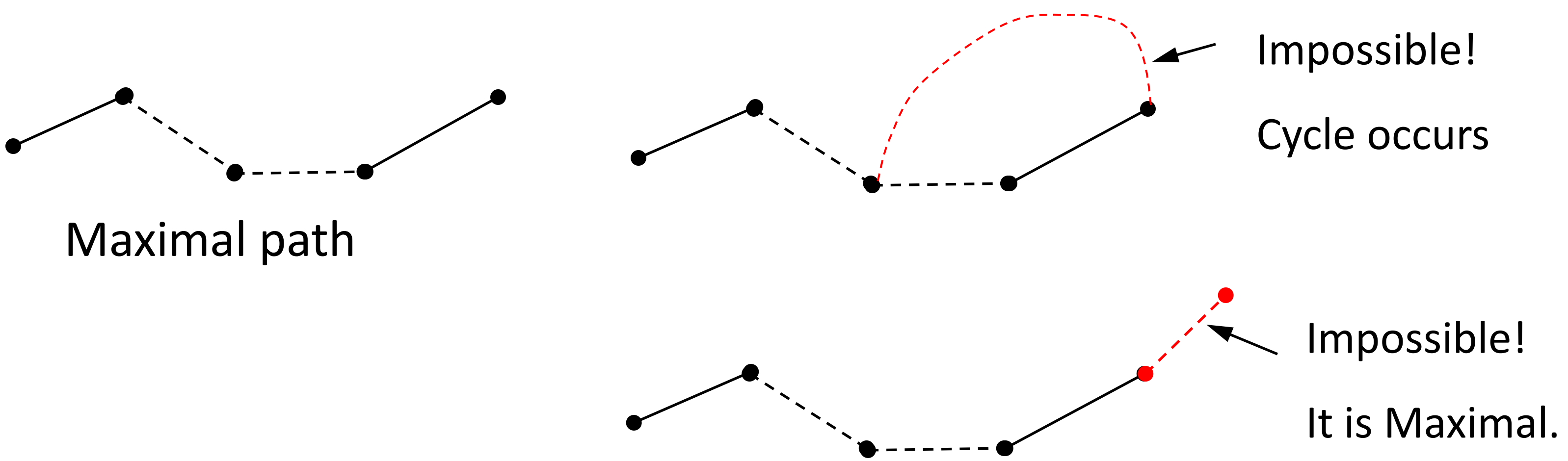
- **Example:**
- $|T_2| = 2^{2-2} = 1$, $|T_3| = 3^{3-2} = 3$, $|T_4| = 4^{4-2} = 16$

Example of T4



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.



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. ■

Discussions:

- **Lemma 2.1.3** implies that every tree with more than one vertex arises from a smaller tree by adding a vertex of degree 1 (all our graphs are finite).

- The next slide is on the proof of equivalence of characterizations of trees using induction, prior results, a counting argument, extremality, and contradiction.

Characterization of Trees

Theorem 2.1.4. For an n-vertex graph G (with $n \geq 1$), the following 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 2.1.4 *Continue*

- A: G is connected and has no cycles
- A⇒{B,C}. **connected, acyclic ⇒ n-1 edges**

  We use **induction** on n.

  - For n=1, an acyclic 1-vertex graph has no edge.
  - For n>1, we suppose that implication holds for graphs with fewer than n vertices.
  - Given an acyclic connected graph G, Lemma 2.1.3 provides a leaf v and states that $G'=G-v$ also is acyclic and connected.
  - Applying the induction hypothesis to G' yields $e(G')=n-2$.
  - Since only one edge is incident to v,

    we have $e(G)=e(G')+1=(n-2)+1=n-1$.

Theorem 2.1.4 continue

B: G is connected and has n-1 edges

- $B \Rightarrow \{A, C\}$. **connected and n-1 edges $\Rightarrow$ acyclic**
  - If G is not acyclic, 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 (by Theorem 1.2.14).
  - From above implications, it implies that $e(G')=n-1$.
  - Since we are given $e(G)=n-1$, no edges were deleted.
  - Thus $G'=G$, and G is acyclic.

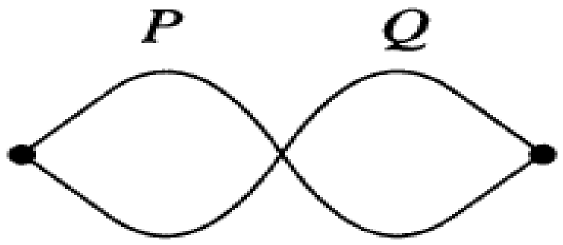
Theorem 2.1.4 continue

C: G has n-1 edges and no cycles

- C⇒{A, B}. **n-1 edges and no cycles ⇒ connected**
  - Let $G_1,\ldots,G_k$ be the components of G.
  - Since every vertex appears in one component, $\Sigma_i n(G_i)=n$.
  - Since G has no cycles, each component satisfies property A. Thus $e(G_i)=n(G_i) - 1$.
  - Summing over i yields $e(G)=\Sigma_i[n(G_i)-1]=n-k$.
  - We are given $e(G)=n-1$, so $k=1$, and G is connected.

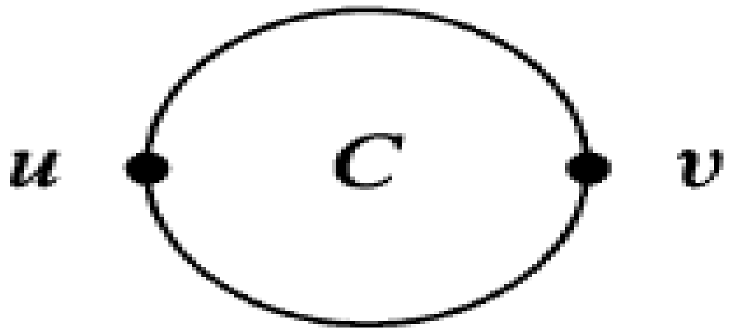
Theorem 2.1.4

- $A \Rightarrow D$. **Connected and no cycles $\Rightarrow$ For *u, v*$\in$*V*(*G*), one and only one *u, v*-path exists.**
  - Since G is connected, each pair of vertices is connected by a path.
  - If some pair is connected by more than one , we choose a shortest (total length) pair P, Q of distinct paths with the same endpoints.
  - By this extremal choice, no internal vertex of P or Q can belong to the other path.
  - This implies that $P \cup Q$ is a cycle, which contradicts the hypothesis A.



Theorem 2.1.4

- $D \Rightarrow A$. **For u, $v \in V(G)$, one and only one u, v-path exists $\Rightarrow$ connected and no cycles.**
  - 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 (this also forbids loops). ■



Corollary: 2.1.5

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

Proof: A tree has no cycles, so Theorem 1.2.14 implies that every edge is a cut-edge.

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

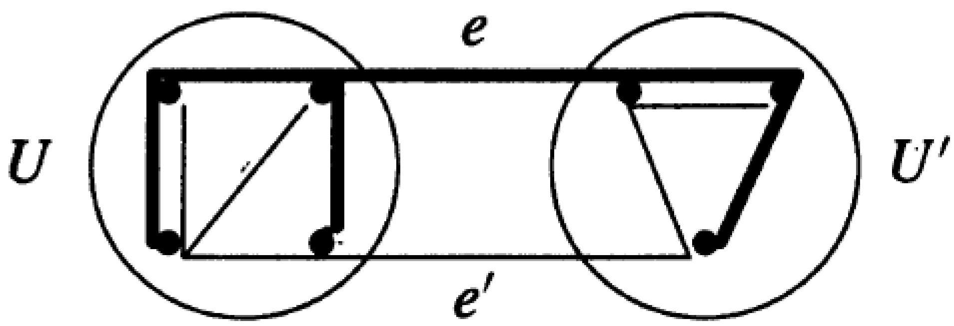
Proof: A tree has a unique path linking each pair of vertices (Theorem 2.1.4D), so joining two vertices by an edge creates exactly one cycle.

c) Every connected graph contains a spanning tree

Proof: As in the proof of B$\Rightarrow$A, C in Theorem 2.1.4, iteratively deleting edges from cycles in a connected graph yields a connected acyclic subgraph. $\blacksquare$

Proposition: If T, T' are spanning trees of a connected graph G and $e \in E(T) - E(T')$, then there is an edge $e' \in E(T') - E(T)$ such that $T - e + e'$ is a spanning tree of G. 2.1.6

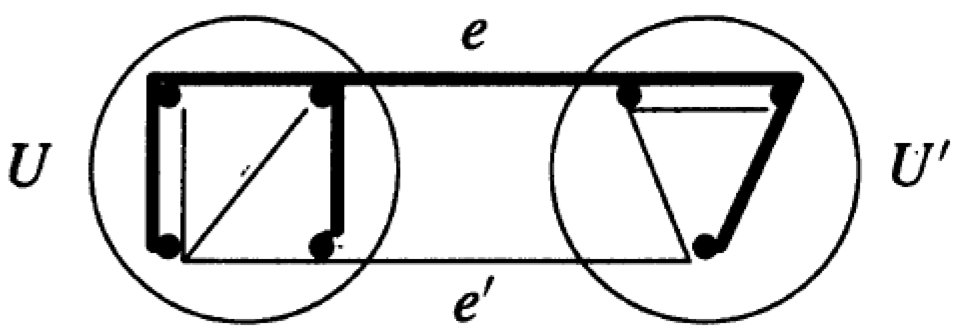
Proof: By Corollary 2.1.5a, every edge of T is a cut-edge of T. Let U and U' be the two components of T-e. Since T' is connected, T' has an edge e' with endpoints in U and U'. Now $T - e + e'$ is connected, has $n(G)$-1 edges, and is a spanning tree of G.



In the figure, T is bold, T' is solid, and they share two edges

Proposition: If T, T' are spanning trees of a connected graph G and $e \in E(T) - E(T')$, then there is an edge $e' \in E(T') - E(T)$ such that $T' + e - e'$ is a spanning tree of G. 2.1.7

Proof: By Corollary 2.1.5b, The graph $T' + e$ contains a unique cycle C. Since T is acyclic, there is an edge $e' \in E(C)-E(T)$. Deleting e' breaks the only cycle in $T' + e$. Now $T' + e - e'$ is connected and acyclic and is a spanning tree of G.



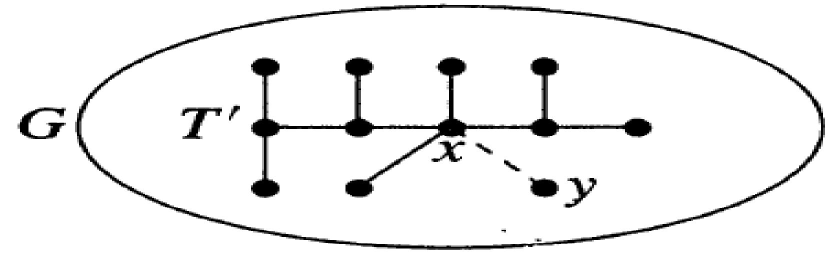
In the figure, adding e to T creates a cycle C of length five; all four edges of $C - e$ belong to $E(T)-E(T')$ and can serve as e'.

Proposition: If *T* is a tree with k edges and G is a simple graph with ƺ(G) ≥ k, then T is a subgraph of G.

Proof: We use induction on k.

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

Induction step: k > 0. We assume that the claim holds for trees with fewer than k edges. Since k > 0, Lemma 2.1.3 allow us to choose a leaf v in T; let u be its neighbor. Consider the smaller tree T' = T-v. By the induction hypothesis, G contains T' as a subgraph, since ƺ (G) ≥ k > K-1.



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 $d_G(x)$ ≥ k, x has a neighbor y in G that is not in this copy of T'. Adding the edge xy expands this copy of T' into a copy of T in G, with y playing the role of v.

Distance in trees and Graphs

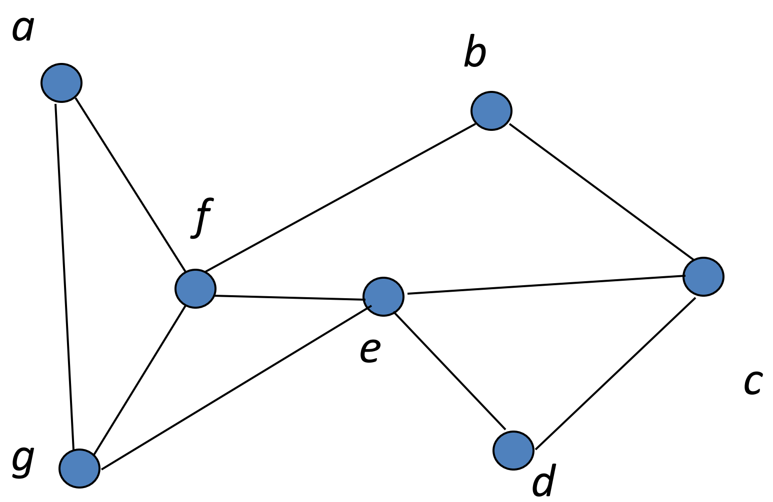
- When using graphs to model communication networks, we want vertices to be close together to avoid communication delays.

- We measure distance using lengths of paths.

Distance in trees and Graphs

- 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** (diam G) is $\max_{u,v\in V(G)} d(u,v)$.
  - Upper bound of 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 *rad* G, is $\min_{u\in V(G)} \varepsilon(u)$.
  - Lower bound of the eccentricity.

Distance, Diameter, Eccentricity, and Radius



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

diameter: 3

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

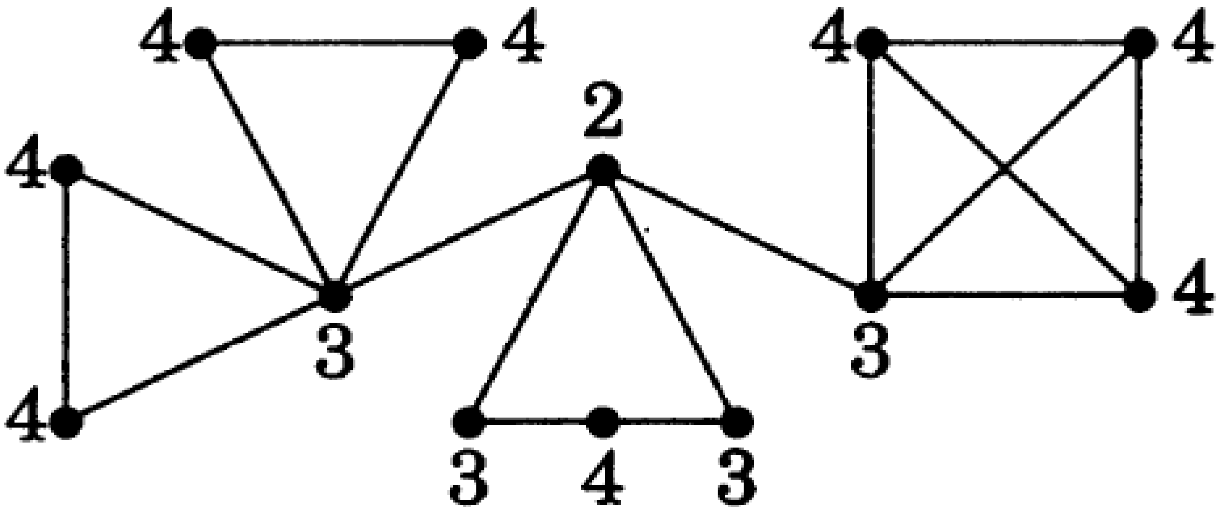
radius: 2

Example

- The **Petersen graph** has diameter 2, since nonadjacent vertices have a common neighbor. The **hypercube** Q_k has diameter *k*, since it takes *k* steps to change all *k*-coordinates.
- The **cycle** C_n has diameter $\lfloor n/2 \rfloor$. In each of these, every vertex has the same eccentricity, and *diam G = rad G*.
- For $n \geq 3$**, the** n**-vertex tree** of least diameter is the star, with diameter 2 and radius 1. The one of largest diameter is the path, with diameter *n-1* and radius $\lceil (n-1)/2 \rceil$. 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.

Example

- 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: If G is a simple graph, then

$diam\ G \geq 3 \Rightarrow diam\ \overline{G} \leq 3$ 2.1.11

Proof:

- When $diam\ G>2$, there exist nonadjacent vertices $u, v \in V(G)$ with no common neighbor

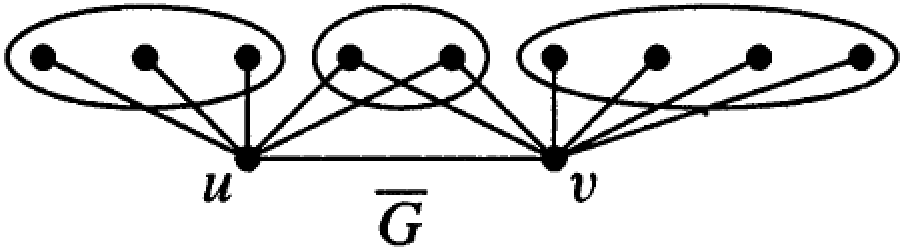
If G is a simple graph, then *diam* $G \geq 3 \Rightarrow$ *diam* $\overline{G} \leq 3$

2.1.11

Proof:

- Hence every $x \in V(G)$-$\{u, v\}$ has at least one of $\{u, v\}$ as a nonneighbor

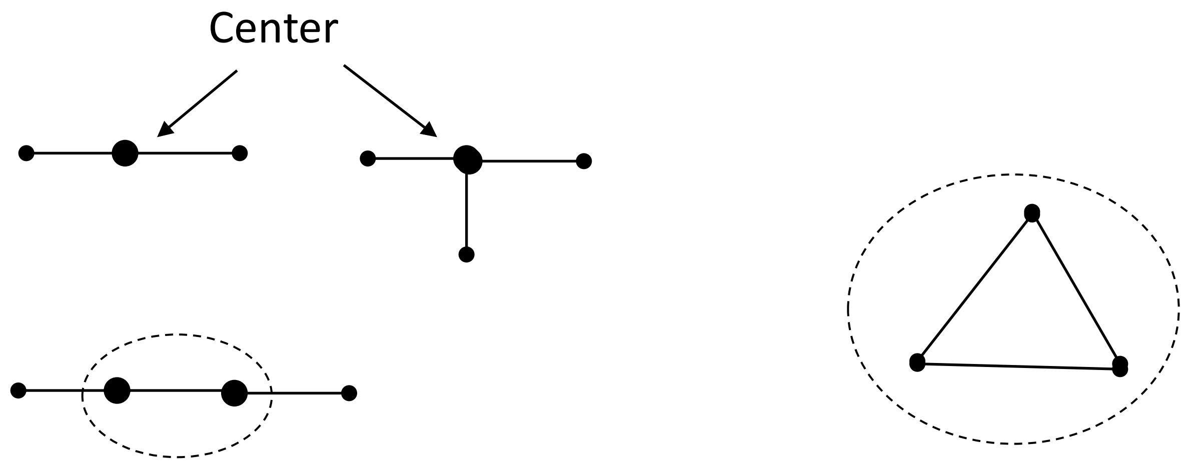
- This makes x adjacent in $\overline{G}$ to at least one of $\{u,v\}$ in $\overline{G}$
- Since also $uv \in E(\overline{G})$, for every pair x, y there is an x, y-path of length at most 3 in $\overline{G}$ through $\{u,v\}$. Hence diam $\overline{G} \leq 3$



Center 2.1.12

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.



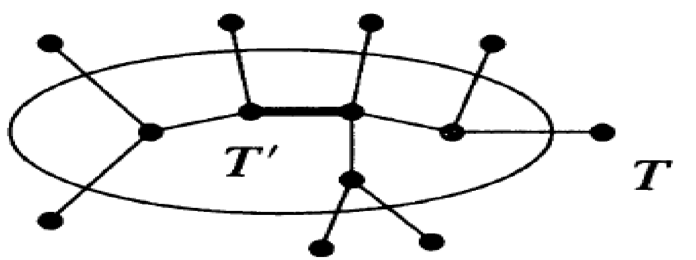
Theorem: The center of a tree is a vertex or an edge 2.1.13

Proof: We use induction on the number of vertices in a tree T.

- **Basis step:** $n(T) \leq 2$. With at most two vertices, the center is the entire tree.

Induction step: $n(T) > 2$

- Form T' by deleting every leaf of T. By Lemma 2.1.3, 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 leave, $\varepsilon_{T'}(u) = \varepsilon_T(u)-1$ 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 $\varepsilon_{T'}(u)$.



- It is shown T and T' have the same center. By the induction hypothesis, the center of T' is a vertex or an edge.

Conclusion

- In this lecture, we have discussed the fundamental properties of trees and the distances in trees and graphs and also discussed the theorems and lemmas based on trees and distance.

- In upcoming lectures, we will discuss spanning trees, enumeration, optimization and trees.