Notes 1 : X-trees

MATH 833 - Fall 2012 *Lecturer: Sebastien Roch*

References: [SS03, Chapter 1,2]

1 Trees

We begin by recalling basic definitions and properties regarding finite trees.

DEF 1.1 (Graph) *A* (finite, undirected, simple) graph $G = (V, E)$ *is an ordered pair consisting of a non-empty finite set V of vertices and a set E of edges each of which is an element of*

$$\{\{x, y\} : x, y \in V,\ x \neq y\}.$$

Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are isomorphic *if there is a bijection $\Psi : V_1 \to V_2$ such that $\{u, v\} \in E_1$ exactly when $\{\Psi(u), \Psi(v)\} \in E_2$. The map Ψ is a* graph isomorphism.

Let $G = (V, E)$ be a graph:

- If $e = \{u, v\} \in E$ then u and v are *neighbors* and e is *incident* with u, v.
- The *degree* $d(v)$ of $v \in V$ is the number of neighbors of v. A vertex of degree 0 is *isolated* and 1 is *pendant*. An edge incident with a pendant vertex is a *pendant edge*.
- We sometimes write $V(G)$ and $E(G)$ for the vertex and edge sets of G. A graph H is a *subgraph* of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$.

DEF 1.2 (Connectedness) *A* (simple) path *in G is a sequence of distinct vertices $v_1, \ldots, v_k$ such that, for all $i \in \{1, \ldots, k-1\}$, $\{v_i, v_{i+1}\} \in E$. If $\{v_k, v_1\} \in E$ then the subgraph C with vertices $V(C) = \{v_1, \ldots, v_k\}$ and edges $E(C) = \{v_1, v_2\} \cup \cdots \cup \{v_k, v_1\}$ is a* cycle. *A graph $G = (V, E)$ is* connected *if for all $u, v \in V$, there is a path between u and v. If G is not connected, the maximal connected subgraphs of G are called its* connected components.

DEF 1.3 (Tree) *A* forest *is a cycle-free graph. A* tree $T = (V, E)$ *is a connected forest.*

Let $T = (V, E)$ be a tree:

- A vertex of T with degree 1 is called a *leaf*. All other vertices of T are *interior* vertices.
- An edge of T is *interior* if both its end vertices are interior. We denote by $\mathring{V}$ and $\mathring{E}$ the sets of interior vertices and edges.
- A tree is *binary* if all interior vertices have degree 3.

THM 1.4 (Characterization of Trees) *Let $G = (V, E)$ be a graph. Then the following are equivalent:*

1. *G is a tree.*
2. *For all $u_1, u_2 \in V$ there is a unique path between u_1 and u_2.*
3. *G is connected and $|V| = |E| + 1$.*

Before giving the proof, we define graph operations that are useful in induction proofs.

DEF 1.5 (Deletion and Contraction) *Let $G = (V, E)$ be a graph and $v \in V$, $e \in E$.*

- *$G \backslash e$ is the graph obtained from G by* deleting *e.*
- *G/e is the graph obtained from G by* contracting *e (and removing).*
- *$G \backslash v$ is the graph obtained from v by deleting v and all incident edges.*

Proof: The equivalence of 1. and 2. is clear from the connectedness and cycle-freeness of the tree.
1. $\Rightarrow$ 3.) We proceed by induction on the number of vertices. Clearly, 3. is true when $|V| = 1$. Suppose $|V| > 1$. Since the graph is finite and cycle-free, it must be that there is at least one leaf v. Then $G \backslash v = (V', E')$ satisfies 3. and we are done.
1. $\Leftarrow$ 3.) We begin with a lemma.

LEM 1.6 *If $G = (V, E)$ is connected then $|V| \leq |E| + 1$.*

Proof: Clear if $|V| = 1$. Assume $|V| > 1$. Contract any edge and use induction to deduce

$$|V| - 1 \leq |E| - 1 + 1.$$

■

We return to the proof. We proceed by contradiction. Suppose 3. holds but G is not a tree. Then there is an edge e in a cycle and $G \backslash e = (V', E')$ is connected. But then

$$|V| = |V'| \leq |E'| + 1 = |E| - 1 + 1 < |E| + 1,$$

a contradiction.

■

2 X-trees

We come to the fundamental graph-theoretic definition of this course.

DEF 1.7 (X-tree) *An X-*tree *$\mathcal{T} = (T, \phi)$ is an ordered pair where T is a tree and $\phi : X \to V$ is such that X is finite and $\phi(X)$ contains all vertices with degree at most 2. (Note: It is neither surjective nor injective.) Two X-trees $\mathcal{T}_1 = (T_1, \phi_1)$ and $\mathcal{T}_2 = (T_2, \phi_2)$ are* isomorphic *if there is a graph isomorphism Ψ between T_1 and T_2 such that $\phi_2 = \Psi \circ \phi_1$.*

Let $\mathcal{T}$ be an X-tree:

- We sometimes write $T(\mathcal{T})$ and $\phi(\mathcal{T})$ for T and ϕ.
- ϕ is called the *labeling map* of T and T is called the underlying tree.

A special class of X-trees, phylogenetic trees, will be our main object of study. It will become clear when we cover the Splits-Equivalence Theorem and the Tree-Metric Theorem why we need to consider the more general setup of X-trees.

DEF 1.8 (Phylogenetic tree) *A* phylogenetic tree *$\mathcal{T}$ is an X-tree whose labeling map ϕ is a bijection into the leaves of its underlying tree T. $\mathcal{T}$ is* binary *if all interior vertices of T have degree 3. We denote by $B(n)$ the set of all binary phylogenetic trees where $|X| = n$. Unless stated otherwise, we let $X = \{1, \ldots, n\} \equiv [n]$.*

THM 1.9 *Every $\mathcal{T} \in B(n)$ has n pendant edges, $n - 3$ interior edges and $n - 2$ interior vertices.*

Proof: Summing over the degrees amounts to counting each edge twice so that

$$1 \cdot (|V| - |\mathring{V}|) + 3 \cdot (|\mathring{V}|) = 2|E|,$$

which implies

$$n + 3(|V| - n) = 1 \cdot (|V| - |\mathring{V}|) + 3 \cdot (|\mathring{V}|) = 2|E| = 2|V| - 2,$$

by Theorem 1.4 and the fact that there are n leaves. Hence $|V| = 2n - 2$, $|E| = 2n - 3$, and $|\mathring{V}| = n - 2$. ∎

THM 1.10 (Counting phylogenetic trees) *Letting $b(n) = |B(n)|$, for all $n \geq 3$*

$$b(n) = 1 \times 3 \times \cdots \times (2n - 5) \equiv (2n - 5)!! \sim \frac{1}{2\sqrt{2}} \left(\frac{2}{e}\right)^n n^{n-2}.$$

Proof: We proceed by induction. The result is clear for $n = 3$. Assume $n > 3$. Consider the map from $B(n)$ to $B(n - 1)$ which removes leaf n and its incident edge and suppresses the resulting degree 2 node. The result follows from Theorem 1.4. The asymptotic formula is obtained from Stirling's formula,

$$n! \sim \sqrt{2\pi} e^{-n} n^{n+1/2}.$$

(Recall that $f(n) \sim g(n)$ indicates $\lim_{n\to\infty} f(n)/g(n) = 1$.) ∎

3 Rooted X-trees

DEF 1.11 (Rooted trees) *A* rooted tree *is a tree $T = (V, E)$ with a single distinguished vertex ρ. A* rooted X-tree *is an X-tree $\mathcal{T} = (T; \phi)$ whose tree T is rooted and whose labeling map ϕ is such that $v \in \phi(X)$ for all $v \in V - \{\rho\}$ of degree at most 2. A* rooted phylogenetic tree *is a phylogenetic tree $\mathcal{T} = (T, \phi)$ whose root has degree at least 2. A* rooted binary phylogenetic tree *is a rooted phylogenetic tree such that every interior vertex has degree 3 except the root which has degree 2. We denote by $RB(n)$ the set of all rooted binary phylogenetic trees where $|X| = n$.*

Recall that a partial order on a set S is a relation $\leq$ such that for all $x, y, z \in S$:

1. (Reflexivity) $x \leq x$.

2. (Antisymmetry) If $x \leq y$ and $y \leq x$ then $x = y$.

3. (Transitivity) If $x \leq y$ and $y \leq z$ then $x \leq z$.

Let $\mathcal{T} = (T, \phi)$ be a rooted X-tree:

- A partial order $\leq_T$ on the vertex set V of T is obtained by letting $v_1 \leq_T v_2$ if the path between the root ρ and v_2 goes through v_1. We say that v_1 is an *ancestor* of v_2 and v_2 is a *descendant* of v_1.

- The *most recent common ancestor (MRCA)* of $A \subseteq X$ is the greatest lower bound of $\phi(A)$ under $\leq_T$.

THM 1.12 (Counting rooted trees) *For all* $n \geq 3$

$$|RB(n)| = (2n-3)!!$$

Proof: There is a natural bijection between $B(n+1)$ and $RB(n)$: remove leaf $n+1$ and the edge incident to it, and root the tree at the degree 2 node so created. ■

Further reading

The definitions and results discussed here were taken from Chapter 2 of [SS03]. Much more on the subject can be found in that excellent monograph. See also [SS03] for the relevant bibliographic references.

References

[SS03] Charles Semple and Mike Steel. *Phylogenetics*, volume 24 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2003.