Lecture 1: Graphs

Cuong Van Than

April 10, 2014

We give a brief introduction to basic concepts in graph theory that are essential to phylogenetics.

Simple graphs. A simple graph G is an ordered pair $(V(G), E(G))$ consisting of a non-empty set $V(G)$ of vertices and $E(G)$ of edges. If each element of $E(G)$ is a pair $\{u, v\}$, where u, $v \in V(G)$, then we say that G is a (simple) undirected graph. If each element of $E(G)$ is an *ordered* pair (u, v), where u, $v \in V(G)$, then we say that G is (simple) directed graph. We will restrict ourselves to simple graphs in this course.

A graph can be represented graphically with vertices indicated by points and edges by lines joining their end vertices; if G is directed, then arrows can be added to indicate the direction of the edges. Note that a graph can have more than one graphical representation. Concepts such as incidence and adjacency are suggested by the graphical representation. An edge is said to be incident to its end vertices, and vice versa. The end vertices are said to be adjacent.

Let G be an undirected graph, and let v be a vertex of G. The neighborhood of v is the set $N(v) = \{w \mid v \text{ and } w \text{ are adjacent}\}$. The degree of v, denoted by $\deg(v)$, is the size (i.e. the number of elements) of $N(v)$. Clearly, $0 \leq \deg(v) \leq |V(G)| - 1$ for any v. If $\deg(v) = 0$, we call v an isolated vertex. If $\deg(v) = 1$, we say v is a pendant vertex (or leaf), and call the edge incident to v a pedant edge.

Theorem 1. *Let G be a simple undirected graph. Then*

$$\sum_{v \in V(G)} \deg(v) = 2|E(G)|.$$

Proof. The theorem follows from the fact that each edge of G is incident to exactly two vertices. $\square$

Corollary 2. *The number of vertices of a simple undirected graph G that have odd degrees is even.*

We have similar definitions for directed graphs. We denote by $N^+(v)$ and $N^-(v)$ the sets $\{w \mid (v, w) \in E(G)\}$ and $\{u \mid (u, v) \in E(G)\}$, respectively, and by $\deg^+(v) =$

$|N^+(v)|$ and $\deg^-(v) = |N^-(v)|$, respectively. If $\deg^+(v) = 0$, then we call it a sink; if $\deg^-(v) = 0$, we call it a source.

Theorem 3. *Let G be a simple directed graph. Then*

$$\sum_{v \in V(G)} \deg^+(v) = \sum_{v \in V(G)} \deg^-(v) = |E(G)|.$$

Graph isomorphism. As noted before, a graph can have more than one graphical representation. Thus it is plausible that two graphical representations that look different in fact represent the same graph structure. Let G and H be two simple directed graphs. We say that G and H are isomorphic if there exists a bijective mapping $\phi\colon V(G) \to V(H)$ such that $\{u, v\} \in E(G)$ if and only if $\{\phi(u), \phi(v)\} \in E(H)$. In other words, the bijective mapping ϕ preserves vertex adjacency. We clearly have a similar definition for isomorphic simple directed graphs.

Up to graph isomorphism, there can be a finite number of graphs with a fixed number of vertices. We introduce some special graphs below. A graph in which each vertex is adjacent to every other vertex is called a complete graph, and is denoted by K_n (n is the number of vertices). The complement of K_n is the graph consisting of n isolated vertices. (The complement of G, denoted by $\overline{G}$, is the graph in which $\{u, v\}$ is an edge of $\overline{G}$ precisely if $\{u, v\}$ is not an edge of G, and vice versa.) A connected graph in which every vertex has degree two is called a cycle graph C_n. The graph obtained from C_n by removing one edge is called the path graph P_n. If the vertex set of a graph can be partitioned into two (nonempty) subsets X and Y such that every edge of the graph is incident to exactly one element of X and exactly one element of Y, then we say G is bipartite. A complete bipartite graph is a bipartite graph in which each element of X is adjacent to every element of Y, and we denote it by $K_{m,n}$, where $m = |X|$ and $n = |Y|$.

Subgraphs and induced subgraphs. A graph H is a subgraph of G if $V(H)$ and $E(H)$ are subsets of $V(G)$ and $E(G)$, respectively. We say that H is an *induced* subgraph of G if $E(H)$ consists of edges of G that have both end vertices in $V(H)$. The subgraph of G induced by a subset $V' \subset V(G)$ is denoted by $G[V']$. The subgraph of G induced by $V(G) \setminus V'$ (i.e. $G[V(G) \setminus V']$) is also denoted by $G - V'$. In the case V' is a singleton set $\{v\}$, we write $G - v$ for $G - \{v\}$.

We note that a subgraph of G can be obtained by deleting certain vertices (along with edges incident to them) and/or edges. If only the removal of vertices from G is allowed, the resulting graph is an induced subgraph of G. This interpretation justifies the notation $G - V'$ above.

We say that H is a spanning subgraph of G if $V(H) = V(G)$ and $E(H) \subset E(G)$. A clique of G is a subgraph of G that is also a complete graph. A subset S of $V(G)$ is called an independent set if the subgraph of G induced by S has no edge.

Walks, trails, and paths. A walk in G is a sequence $W = v_0 e_1 v_1 \dots e_k v_k$ of alternating vertices and edges such that for $1 \leq i \leq k$, the end vertices of e_i are v_{i-1} and v_i. If the edges of W are distinct, we call W a trail. If the vertices of W are distinct, we call W a path, and since we assume G is simple, we can denote a path by the sequence of vertices in it.

If the first and last vertices of W are identical, we say that W is closed. A cycle is a closed walk in which all the vertices, except for the first and last, are distinct.

Theorem 4. *Let W be a walk from u to v. Then W contains a path from u to v.*

Two vertices u and v of G are said to be connected if there exists a path $v_0, v_1, \dots, v_k$ in G with $v_0 = u$ and $v_k = v$. It is easy to see that the relation "being connected" is an equivalence relation on $V(G)$. Thus $V(G)$ can be partitioned into equivalence classes $V_1, V_2, \dots, V_m$. The induced subgraphs $G[V_1], G[V_2], \dots, G[V_m]$ are called the connected components of G. If G has only one connected component, we say that G is connected.

Theorem 5. *Let G be a simple undirected graph with n vertices and k edges. Then G has at least $n - k$ connected components.*

The length of a walk is defined as the number of edges in it. We say a walk is odd/even if it has odd/even length. The following theorem characterizes bipartite graphs in terms of odd cycles.

Theorem 6. *A simple undirected graph G is bipartite if and only if it has no cycle of odd length.*

Proof. Suppose first that G is bipartite. Then $V(G)$ can be partitioned into two subsets X and Y such that every edge of G is incident to exactly one vertex in X and exactly one vertex in Y. Let $C = v_1, \dots, v_{k-1}, v_k = v_1$ be a cycle. As the vertices in C alternate between X and Y, it must be that k is an odd number, and hence the length of C is even.

Conversely, suppose that G has no cycle of odd length. We can assume that G is connected, for otherwise we just apply the following construction to each connected component of G. Choose a vertex $u \in V(G)$, and for every $v \in V(G)$ let $d(u, v)$ be the distance from u to v (i.e. the length of a shortest path from u to v). Let X be the set of vertices whose distance from u is even, and let Y be the set of vertices whose distance from u is odd. Except for the case G is a trivial graph consisting of a single vertex, both X and Y are nonempty (X contains u, and Y contains $N(u)$). Assume that there is an edge between v and w, both of which are in X. We then have a closed walk W consisting of a shortest path from u to v, edge $\{v, w\}$, and a shortest path from w to u. Clearly W has odd length, and hence it contains an odd cycle (see Problem 4). This contradicts our assumption. Similarly, we would obtain a contradiction if there were an edge between two elements of Y. The claim now follows. $\square$