Lecture 3: Split Networks

Cuong Van Than

April 24, 2014

We introduced the concept of split compatibility, and showed that a set Σ of pairwise compatible X-splits uniquely determines an X-tree T that induces exactly the X-splits of Σ (i.e. $\Sigma(T) = \Sigma$). We can explicitly reconstruct the tree T by employing the following theorem.

Theorem 1. *Let (T, ϕ) be an X-tree, and let $A|B$ be an X-split such that it is not in $\Sigma(T)$, but it is compatible with every X-split of $\Sigma(T)$. Then there exists a unique vertex v of T such that for each component (V, E) of $T - v$, $\phi^{-1}(V)$ is a subset of either A or B.*

The theorem provides us with a recursive procedure for computing an X-tree that induces a given set Σ of compatible X-splits. Let T' be the X-tree constructed for $\Sigma' = \Sigma \setminus \{A|B\}$, and let v be vertex determined as in Theorem 1. We then "split" v into two vertices v_A and v_B, and reattach the components of $T - v$ whose vertex-label sets are subsets of A to v_A, and those whose vertex-label sets are subsets of B to v_B. The resulting tree clearly induces all the X-splits in Σ.

The situation is a bit more complicated when there are incompatible X-splits in Σ, and we introduce in this lecture split networks as a generalization of X-trees that can represent incompatible X-splits.

Let $G = (V, E)$ be an undirected (simple) graph. A k-edge coloring κ of G is an assignment of k colors to the edges of G ("colors" is used in the abstract sense—colors can be numbers $1, \ldots, k$, class names, or X-splits as in this lecture), that is, κ is a function from E to a set of colors K. We assume throughout the lecture that all k-edge colorings are surjective, meaning that each color is assigned to at least one edge of G (or equivalently, $\kappa(E) = K$). A coloring κ is proper if no two adjacent edges have the same color.

We say that a proper coloring κ is isometric if for every pair of vertices u and v, the edges along any shortest path joining u and v all have distinct colors, and the set of colors of the edges along any shortest path is the same. More formally, let $d(u, v)$ be the distance between u and v (i.e. the number of edges in a shortest path joining u and v), and let $C_\kappa(P)$ be the set of colors of the edges in P. If for every two vertices u and v, and for any two shortest paths P and and P' connecting u and v, we always have $C_\kappa(P) = C_\kappa(P')$ and $|C_\kappa(P)| = |C_\kappa(P')| = d(u, v)$, then we say that κ is isometric.

A split graph is a connected bipartite graph $G = (V, E)$ together with a (surjective) isometric edge coloring κ.

Theorem 2. *Let G be a split graph (with an isometric edge coloring κ). Then for any color c in the range of κ, the deletion of the edges of G that are assigned color c results in exactyl two connected components.*

Proof. Denote by $C_\kappa(u, v)$ the set of colors of edges in a shortest path joining u and v (it is well-defined, for κ is isometric). We define two vertex sets

$$X = \{x \in V(G) \mid c \notin C_\kappa(u, x)\} \quad \text{and} \quad Y = \{y \in V(G) \mid c \notin C_\kappa(v, y)\},$$

and claim that the two subgraphs of G induced on X and Y are the two (and the only two) components of the graph G with all the c-colored edges removed.

Consider a vertex w of G. Then $d(u,w)$ and $d(v,w)$ can differ by at most one (why?). Further, they are not equal, for otherwise G would contain an odd cycle. We first assume that $d(u,w) = d(v,w) + 1$. This implies that if P is a shortest path from v to w, then path P augmented with edge (u,v) is a shortest path from u to w. Since κ is isometric, no edges in P have color c, and hence $w \in Y$. Assume now that $d(v,w) = d(u,w) + 1$, then the same arguments we have $w \in X$. Thus for any vertex w of G, it is in either the graph $G[X]$ or $G[Y]$. The two graphs $G[X]$ and $G[Y]$ are both connected, and hence the removal of all the c-colored edges of G results in at most two connected components.

Suppose that there exists a vertex w that belongs to both X and Y. As noted previously, $d(u,w) = d(v,w) + 1$ or $d(v,w) = d(u,w) + 1$. In either case, we have two shortest paths joining two vertices (u and w for the first case, and v and w for the second case) whose edge color sets are different (one of them contains color c, the color of edge (u,v)), which contradicts that κ is isometric. Hence, X and Y form a partition of the vertex set of G. Further, any edge joining a vertex in X and a vertex in Y must have color c (why?). It then follows that $G[X]$ and $G[Y]$ are the two connected components of the graph G with all the c-colored edges removed. $\square$

Let Σ be a set of X-splits. A split network is a split graph $G = (V, E)$ with an isometric coloring $\kappa\colon E \to \Sigma$ and a vertex labeling $\phi\colon X \to V$ so that for every X-split $A|B \in \Sigma$, the deletion of the edges of G that are assigned color $A|B$ results in two connected components the label sets of whose vertices are exactly A and B.

Observe that the definition in the preceding paragraph gives no guarantees that a split network exists for every set Σ of X-splits. Of course, we know that every set Σ of pairwise compatible X-splits uniquely determines an X-tree, which is a split network. We will consider canonical split networks (also called Buneman graphs) that represent incompatible X-splits.