

## Building phylogenetic networks from triplet data

Leen Stougie (Free University and CWI Amsterdam) 13/11/2008.

Phylogenetics is the field at the interface of biology, mathematics and computer-science which studies the (re-)construction of plausible evolutionary scenarios when confronted with incomplete and/or error-prone biological data. Until recently almost all research effort was directed at finding evolutionary trees. However, biological phenomena such as hybridisation, horizontal gene transfer, recombination, and gene duplication, especially occurring frequently in lower order species, cause evolution not to be representable by a tree structure.

In this lecture we will concentrate on building phylogenetic networks, rather than trees, from so-called triplet data. A triplet represents a rooted phylogenetic tree on a subset of three species. Given a family of such triplets, we like to find a rooted network consistent with all triplets. Such a network has a root, and leaves (the species) and otherwise splitvertices with indegree one and outdegree two, and reticulation vertices with indegree two and outdegree one. Preferably we wish the network to have a low level of nestedness. A level- $k$  network is a network of which every biconnected component contains at most  $k$  reticulation vertices. In general, given a triplet set, knowing if there exists a consistent network of level at most  $k$  is an NP-complete problem, for any  $k$  greater than or equal to 1. However, as we will show, the problem can be solved in polynomial time for  $k = 1$  and  $k = 2$  if the triplet set is dense, meaning that for every three species there exist at least one triplet.

We also discuss related problems, like minimizing the number of reticulation vertices or maximising the number of consistent triplets, and finish with a bag of intriguing open problems in this field.