Workshop on Foundations of Phylogenetic Networks
6–8 September 2023

Abstracts

Talk 1

Identifiability of species networks from multilocus data under the Network Multispecies Coalescent model

Elizabeth S. Allman, University of Alaska Fairbanks, USA

Abstract

This talk presents several recent advances in our understanding of which network features can be identified from summary data under the Network Multispecies Coalescent model. Multilocus datasets might be summarized by pairwise log-det distances between genomic sequences, or by quartet concordance factors from inferred gene trees. We discuss topological network parameter identifiability from these data, emphasizing distance-based results. We then discuss practical implementations and algorithms built on these ideas, and outline future research directions.
This is joint work with J. Rhodes, H. Banos and others.

Talk 2

Recovering the species tree under lateral gene transfer

Sebastien Roch, University of Wisconsin-Madison, USA

Abstract

Lateral gene transfer (LGT) is a common cause of gene tree incongruence in phylogenomic analyses. I will discuss some recent results on the reconstruction of the species tree under a stochastic model of LGT.

Talk 3

The emergence of branching structures in models of reticulated evolution

François Bienvenu, ETH Zürich, Switzerland

Abstract

Even when evolution is highly reticulated, in many settings we can expect phylogenetic networks to have a tree-like structure on the large scale. In this talk, I will explain how this idea can be formalized and present a few models that illustrate the challenges associated to the identification of such branching structures in phylogenetic networks. This is joint work with Jean-Jil Duchamps.

Talk 4
A branching process approach to level-k phylogenetic networks
Benedikt Stufler, Vienna University of Technology, Austria

Abstract
The mathematical analysis of random phylogenetic networks via analytic and algorithmic methods has received increasing attention in the past years. In the present talk we introduce branching process methods to their study. Our main results focus on random level-k networks with n labeled leaves. Although the number of reticulation vertices in such networks is typically linear in n, we prove that their asymptotic global and local shape is tree-like in a well-defined sense. We show that the depth process of vertices in a large network converges towards a Brownian excursion after rescaling by the square root of n. We also establish Benjamini–Schramm convergence of large random level-k networks towards a novel random infinite network.

Talk 5
Limit laws for patterns in ranked tree-child networks
Michael Fuchs, National Chengchi University, Taiwan

Abstract
Ranked tree-child networks were recently introduced by Bienvenue, Lambert, and Steel who proved that the number of cherries of a random ranked tree-child network becomes the Poisson law as the number of leaves tend to infinity. This is, up to our knowledge, one of the first examples of a limit law result for a pattern in random phylogenetic networks. In this talk, we will explain how limit laws for other patterns in random ranked tree-child networks can be derived. Moreover, for a general pattern, we will present a conjecture which says that the limit law can be only one of three types.

This is joint work with Hexuan Liu and Tsan-Cheng Yu

Talk 6
A Database Approach to Solve the Tree Containment Problem in Phylogenetic Networks
Philippe Gambette, Université Gustave Eiffel, France

Abstract
The NP-hard Tree Containment problem consists in deciding whether a phylogenetic tree is displayed by a phylogenetic network. Considering this problem as a special case of directed subgraph homeomorphism allows us to design several algorithmic

approaches, some inspired by database theory where the network is considered as the database and the tree as the query, to solve it efficiently in practice. We compare those approaches on randomly generated networks and trees, as well as on data from the PhyloCOG dataset. In particular, we explore the question of whether phylogenetic networks reconstructed from the clusters of input rooted trees are actually likely to display those trees or not.

This is joint work with Sarah J. Berkemer (École Polytechnique), Pierre Bourhis (CNRS), Lionel Seinturier (Université de Lille) and Marion Tommasi (Inria)

Talk 7
Extended µ-representation for the identification and comparison of orchard phylogenetic networks
Joan Carles Pons, University of the Balearic Islands, Spain

Abstract
Although the comparison of phylogenetic trees is well studied, and there are multiple ways to do it in an efficient way, the situation is much different for phylogenetic networks. Some classes of phylogenetic networks, mainly tree-child networks, are known to be classified efficiently by their µ-representation, which essentially counts, for every node, the number of paths to each leaf. In this talk we explain the extended µ-representation of networks, where the number of paths to reticulations is also taken into account. This modification allows us to distinguish orchard networks and to define a sound metric on the space of such networks that can, moreover, be computed efficiently.

This is joint work with Gabriel Cardona, Gerard Ribas and Tomás Martínez (UIB)

Talk 8
Labellable phylogenetic networks
Andrew Francis, Western Sydney University, Australia

Abstract
The set of forests of phylogenetic trees is in bijection with the set of partitions of finite sets, but is there such a set-theoretic bijection for phylogenetic networks? In this talk I will describe such a bijection between a large class of networks, called "labellable", and sets of covers of finite sets that are "expanding". This new class of networks contains most well-studied classes, such as the orchard and the tree-sibling classes, but does

not contain all of the tree-based networks. It can be structurally characterised in graph-theoretic terms.

This is joint work with Peter Jarvis (Tasmania) and with Mike Steel (Canterbury).

Talk 9
Model selection for phylogenetic networks data with cross-validation
James Degnan, University of New Mexico, USA

Abstract
Topological phylogenetic networks at the species level can be inferred from gene trees, evolutionary trees estimated at different loci. Several statistical inference methods have been used under the multispecies coalescent model of a species network. We examine a rooted triple version of cross-validation to distinguish between candidate networks and compare with using full gene trees.

Talk 10
Tree of blobs inference under the coalescent model
Hector Banos, California State University, San Bernardino

Abstract
The network multispecies coalescent model (NMSC) describes the formation of gene trees within a species network under the process of incomplete lineage sorting. Inference methods under the NMSC are highly constrained by hefty computational demands, along with the uncertainty of how complicated a network might be and still be consistently inferred. We present a step toward inferring a general species network by showing the identifiability of its tree of blobs. The tree of blobs of a network is obtained after contracting non-cut edges to nodes, so only tree-like relationships between the taxa are shown. We also present TINNIK, a fast, practical algorithm to infer the tree of blobs from a collection of gene trees.

This is joint work with J. Rhodes, E. Allman, and J. Mitchell.

Talk 11
Distinguishing level-1 phylogenetic network-based Markov models
Elizabeth Gross, University of Hawaii at Mānoa, USA

Abstract: One class of phylogenetic models to incorporate reticulate evolutionary events such as hybridization are network-based Markov models. In order for statistical methods for network reconstruction to be consistent using these models, the network parameter needs to be identifiable. Here, we show that the semi-directed network parameter of a triangle-free, level-1 network model with any fixed number of reticulation vertices is generically identifiable under the Jukes–Cantor, Kimura 2-parameter, or Kimura 3-parameter models of DNA substitution.

Talk 12

Network Classes and Their Mathematical Properties

Louxin Zhang, National University of Singapore

Abstract

In the study of algorithmic issues of phylogenetic networks, different classes of phylogenetic networks have been investigated. The talk will cover the mathematical properties of popular network classes, including tree-child networks and galled networks.

Talk 13

Stretched exponentials in the asymptotics of phylogenetic networks

Michael Wallner, TU Wien, Austria

Abstract Let c_n be the number of objects, such as phylogenetic networks, of size n. We are interested in the asymptotics of this sequence, i.e., a "simpler" sequence a_n such that the quotient c_n/a_n converges to 1 (or is bounded) for n to infinity. In particular, we will focus on stretched exponential terms $m^{(n^{(s)})}$ with $m>0$ and $0<s<1$. The presence of such terms is not common, although recently more and more examples emerge. It is generally quite difficult to prove that a sequence has such a stretched exponential, which is partly due to the observation that such a sequence cannot be "very nice", as, e.g., its generating function cannot be algebraic. Previously, the saddle point method was the only generic method for proving such a phenomenon, which, however, requires detailed information on the generating function. Recently, together with Andrew Elvey Price and Wenjie Fang, we have developed a new method on the level of recurrences to prove stretched exponentials. I will introduce the basics of this method and show how we used it together with Yu-Sheng Chang, Michael Fuchs, Hexuan Liu, and Guan-Ru Yu to prove such a phenomenon for d-combining tree-child networks, i.e. networks in which every reticulation node has exactly d parents.

Talk 14

On the Sackin index of galled trees

Bernhard Gittenberger, TU Wien, Austria

Abstract

Galled trees are rooted phylogenetic networks in which every cycle has exactly one tree vertex above all other vertices of that cycle and exactly one reticulation vertex, which must then be below all its vertices. We consider galled trees with labeled leaves and their Sackin index, defined as the sum over all depths of the leaves. We use a combinatorial specification to describe the class of networks and derive functional equations for their generating functions. The analysis of the Sackin index is then done by studying the singularity structure of these generating functions.
This is joint work with Michael Fuchs.

Talk 15

On the joint distribution of the numbers of cherries and pitchforks for the Ford model

Kwok Pui Choi, National University of Singapore, Singapore

Abstract

Distributional properties of tree shape statistics under random phylogenetic tree models play an important role in investigating the evolutionary forces underlying the observed phylogenies. In this talk, we focus on two subtree counting statistics, the number of cherries and the number of pitchforks for a tree with leaves generated by the Ford model. Ford model is a one-parameter family of random phylogenetic tree models which includes the proportional to distinguishable arrangement (PDA) and the Yule models, two tree models commonly used in phylogenetics. Based on a non-uniform version of the extended Polya urn models in which negative entries are permitted for their replacement matrices, we obtain the strong law of large numbers and the central limit theorem for the joint distribution of and for the Ford model. Furthermore, we derive a recursive formula for computing the exact joint distribution of and . This leads to exact formulas for their means and higher order asymptotic expansions of their second moments, which allows us to identify a critical parameter value for the correlation between and . That is, when the number of tree leaves is sufficient, and are negatively correlated for alpha in [0, ½]; and positively correlated for alpha in (½, 1). This talk is based on a joint work with Gursharn Kaur and Taoyang Wu.

Talk 16

Bijections for phylogenetic networks

Guan-Ru Yu, National Kaohsiung Normal University, Taiwan

Abstract

Bijective proofs are widely used in combinatorial counting problems. In this talk, I will explain the role bijective proofs have recently played in the counting of phylogenetic networks (tree-child networks and ranked tree-child networks).

The talk is based on joint work with Alessandra Caraceni (Scuola Normale Superiore), Michael Fuchs (National Chengchi University) and Louxin Zhang (National Singapore University).

Talk 17

Finding redundancy in phylogenetic networks

Charles Semple, University of Canterbury, New Zealand

Abstract

Examining the multi-set of phylogenetic trees embedded in a phylogenetic network arises in many recently studied problems such as the Tree-Containment problem and counting the number of phylogenetic trees displayed by a network. The size of this multi-set is exponential in the number of reticulation vertices. Thus it is natural to ask whether we can remove a reticulation vertex so that the resulting network still embeds the same set of trees, but the size of the multiset of embedded phylogenetic trees is much less. In this talk, we investigate this question.
This is joint work with Simone Linz.

Talk 18

How far is my network from being edge-based? Proximity measures for edge-basedness of unrooted phylogenetic networks

Mareike Fischer, University of Greifswald, Germany

Abstract

Phylogenetic networks which are, as opposed to trees, suitable to describe processes like hybridization and horizontal gene transfer, play a substantial role in evolutionary research. However, while non-treelike events need to be taken into account, they are relatively rare, which implies that biologically relevant networks are often assumed to be

similar to trees in the sense that they can be obtained by taking a tree and adding some additional edges. This observation led to the concept of so-called tree-based networks, which recently gained substantial interest in the literature. Unfortunately, though, identifying such networks in the unrooted case is an NP-complete problem. Therefore, classes of networks for which tree-basedness can be guaranteed are of the utmost interest. The most prominent such class is formed by so-called edge-based networks, which have a close relationship to generalized series parallel graphs known from graph theory. They can be identified in linear time and are in some regards biologically more plausible than general tree-based networks. While concerning the latter proximity measures for general networks have already been introduced, such measures are not yet available for edge-basedness. This means that for an arbitrary unrooted network, the "distance" to the nearest edge-based network could so far not be determined. In my talk, I will fill this gap by introducing two classes of proximity measures for edge-basedness.

(This project is joint work with Tom Hamann and Kristina Wicke.)

Talk 19
Exploring spaces of semi-directed phylogenetic networks

Kristina Wicke, New Jersey Institute of Technology, USA

Abstract
Semi-directed phylogenetic networks have recently emerged as a class of phylogenetic networks sitting between rooted and unrooted phylogenetic networks since they contain directed and undirected edges. For example, software such as PhyloNetworks, NANUQ, and PhyNEST reconstructs semi-directed level-1 networks from biological data. However, in contrast to rooted and unrooted phylogenetic networks, little is known about searching spaces of semi-directed phylogenetic networks to find an optimal network. In this talk, we discuss a rearrangement move for semi-directed phylogenetic networks, called cut edge transfer (CET), and show that the space of semi-directed level-1 networks with a fixed leaf set and number of reticulations is connected under CET. By introducing two additional moves that allow for the addition and deletion of reticulations, we extend our results to semi-directed (level-1) networks on a fixed leaf set. As a byproduct, we also obtain connectedness results for rooted level-1 networks under a rooted version of CET.

This is joint work with Simone Linz (University of Auckland)