

# Relazione attività RTDb di Filippo Disanto (II anno)

---

L'attività svolta durante il secondo anno di servizio come RTDb presso il Dipartimento di Matematica è riassunta qui di seguito.

## (1) Didattica svolta nel periodo in esame

- n.40 ore di esercitazione per il corso di Geometria a Fisica (titolare Prof. Mario Salvetti).
- n.42 ore di lezione per il corso di Matematica a Scienze Naturali e Geologia (titolare Prof. Marco Abate).

## (2) Ricerca svolta nel periodo in esame

La mia attività di ricerca si è concentrata sullo studio di strutture combinatorie e modelli probabilistici usati nella descrizione quantitativa di fenomeni biologici. Gli argomenti trattati sono brevemente riassunti qui di seguito. Maggiori dettagli e referenze si trovano nei manoscritti allegati.

- **Revisione di progetti intrapresi nel primo anno di servizio.** Diverse migliorie sono state apportate agli articoli [3.b, 3.c] già in parte oggetto della relazione riguardante il primo anno di servizio. In questi lavori, si sono studiati aspetti matematici e computazionali dell'algoritmo "Neighbor-Joining" per la stima di alberi filogenetici, come anche proprietà enumerative e probabilistiche di strutture combinatorie chiamate "Compact Coalescent Histories" utilizzate in Biologia Computazionale. A seguito di tali modifiche, gli articoli sono al momento in corso di stampa ed accessibili on-line sul sito delle riviste.
- **Alberi filogenetici con massima probabilità: soluzione di un problema aperto.** Nell'articolo [3.a], si risolve in senso affermativo una congettura riguardante la struttura di alberi filogenetici di massima probabilità nel modello "multispecies coalescent". Qui di seguito si riassumono brevemente alcune idee di base.  
Un albero *ordinato* è un albero binario con radice in cui foglie diverse hanno etichette diverse ed i nodi interni sono ordinati linearmente in modo che ogni cammino che dalla radice dell'albero raggiunge una foglia incontri nodi interni in verso crescente (Fig. 1 A). Un albero di *specie* è un albero binario con radice e foglie etichettate in cui i rami hanno una lunghezza misurata in opportune unità. I rami di un albero di specie corrispondono a popolazioni differenti e la struttura dell'albero rappresenta le relazioni di discendenza tra tali popolazioni. Un albero *genetico* è un albero ordinato considerato internamente ad un albero di specie. In Fig. 1, l'albero genetico  $G$  dato in A è realizzato all'interno dell'albero di specie  $S$  dato in B rispettando l'ordine dei nodi interni: nodi di  $G$  con etichetta crescente si dispongono dall'alto verso il basso all'interno di  $S$  determinando un ordine temporale tra le biforazioni di  $G$ . Mentre un albero di specie rappresenta relazioni di discendenza tra popolazioni diverse, un albero genetico interno ad un albero di specie mostra in che modo geni campionati da individui appartenenti alle popolazioni più recenti si sono evoluti nel tempo dentro l'albero di specie.

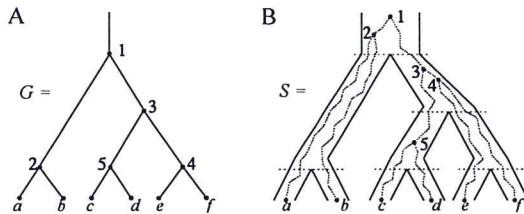


Figura 1: Un albero genetico ed un albero di specie che, a meno dell'ordinamento dei nodi interni, hanno la stessa struttura. (A) Un albero ordinato. (B) Una possibile realizzazione dell'albero ordinato dato in A all'interno di un albero di specie.

Per un dato albero di specie  $S$  con  $n$  foglie, il modello “multispecies coalescent” (Degnan and Rosenberg, Trends Ecol. Evol. 24: 332-340, 2009) permette di definire e calcolare una distribuzione di probabilità sull’insieme degli alberi genetici della stessa taglia di  $S$ . Tale modello stocastico è impiegato per stimare alberi di specie a partire da alberi genetici costruiti prendendo in esame sequenze di DNA appartenenti ad alcuni individui selezionati dalle popolazioni in oggetto. In pratica, considerando molteplici aree genetiche di interesse si crea una collezione di alberi genetici, e successivamente si cercano quegli alberi di specie che massimizzano la probabilità degli alberi genetici osservati.

Usando un approccio esaustivo, Degnan et al. (Math. Biosci. 235: 45-55, 2012; IEEE/ACM Trans. Comput. Biol. Bioinf. 9: 1558-1568, 2012) avevano mostrato che, per un fissato albero di specie con al più 5 foglie, l’albero genetico più probabile ha la stessa struttura dell’albero di specie a meno dell’ordinamento dei nodi interni (Fig. 1 B). Gli stessi autori avevano dunque posto il problema di verificare o smentire tale proprietà per alberi di specie di taglia arbitraria. Nell’allegato [3.a] si fornisce una dimostrazione combinatoria di questo fatto, la cui veridicità ha interessanti conseguenze applicative. In particolare, il risultato ci indica come la struttura non ordinata dell’albero di specie possa essere predetta guardando la struttura degli alberi genetici ordinati che appaiono dai dati con frequenza maggiore. Invece, come mostrato da Degnan and Rosenberg (Plos Genet. 2: 762-768, 2006), guardando alla struttura non ordinata più frequente per gli alberi genetici, si perde la corrispondenza con la struttura non ordinata dell’albero di specie.

### (3) Pubblicazioni e manoscritti a cui ho lavorato nel periodo in esame

- F. Disanto, P. Miglionico, G. Narduzzi. On the unranked topology of maximally probable ranked gene tree topologies, sottomesso a rivista. [Allegato 3.a]
- F. Disanto, N.A. Rosenberg (2018: in press). Enumeration of compact coalescent histories for matching gene trees and species trees, *Journal of Mathematical Biology*. [Allegato 3.b]
- J. Kim, F. Disanto, N.M. Kopelman, N.A. Rosenberg (2018: in press). An extended analysis of the behavior of admixed taxa in the neighbor-joining algorithm, *Bulletin of Mathematical Biology*, Special Issue: Algebraic Methods in Phylogenetics. [Allegato 3.c]

### (4) Conferenze nel periodo in esame

“Enumerative properties of gene tree configurations in matching species trees”, 2nd International Workshop on Enumeration Problems & Applications 2018, Pisa 5-8/11/2018.