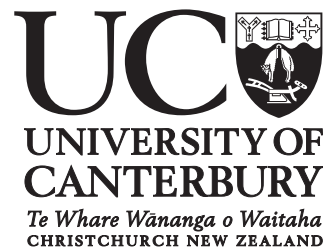# Webs on trees
Understanding ecological network motifs through phylogenetics trees and backward

Giulio Valentino Dalla Riva

Supervisors:

Prof. Mike Steel (Mathematics and Statistics)
Prof. Daniel Stouffer (Biology)

UC
UNIVERSITY OF
CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Research Proposal for the Degree of Doctor of Philosophy
in the Department of Mathematics and Statistics
University of Canterbury

· 2013 ·

Contents

Chapter 1

Introduction

The anthropogenic pressure on the landscape and the living species communities is every day more evident. The biodiversity of the world is recognized as a precious, yet fragile, treasure and more and more efforts are done in order to preserve it, both in the local and in the global scale. A sensible conservation policy requires science to understand the current status of the ecological world and to develop tools with which to forecast its future status.

Biodiversity is, at the same time, the product of fast time processes, such as the one ruling the dynamics of the web of life, and long time processes, such as the ones involved in the growth of the tree of life.

Hence, it is of great importance to understand how this different time scale processes interact.

We approach the problem of studying the interplay between the building blocks of structures of the web and tree of life: motifs. In foodwebs, motifs are given by small subgraphs, which are seen to be over-abundant in the empirical observed webs. In phylogenies we can focus our attention on the mutual evolutionary distance between pair, triplets, quadruplets, ... of species.

In this scenario many elementary question, regarding trophic linking, speciation or extinction events, are yet to be answered.

We will explore the interplay between webs and trees following different approaches, including statistical description, model development and analytical investigation.

Chapter 2

State of the art

Foodwebs are currently analysed in ever growing detail: a number of network characteristics are considered of particular ecological interest - e.g., connectivity, stability, keyroles, degree distribution, motifs distribution. Analogously, phylogenetic trees also topical and their reconstruction is becoming sounder and sounder thanks to the developing sequencing techniques. The relationship between trophic networks and phylogenies is studied within different approaches including statistical descriptions, modelling efforts and analytical studies.

The evolutionary history of a species ensemble has non trivial influence on the species' role in the food webs, as shown by Stouffer et al. [?], Naisbit et al. [?]. Moreover, Eklöf et al. [?] showed that the taxonomy of a species community is a valid predictor of its trophic interactions: the authors find, indeed, a phylogenetic signal in the formation of food webs; similar results are found in Krasnov et al. [?], Ives et al. [?], Bersier et al. [?].

Interpretations of motifs in terms of ecological properties has been reviewed by Bascompte and Stouffer [?], while various field investigations have been and are being attempted (see for example Rip et al. [?]). It has been argued that Phylogenetic Diversity arising from the evolutionary history of a species community may be a predictor of ecosystems functions: see Srivastava et al. [?], Gravel et al. [?], Mouquet et al. [?].

Modelling   A survey of classical models, up to the models developed in 2002, can be found in Drossel and McKane [?]. The author focused on stability issues, covering both static and dynamical models. More recently, a number of models can be found in Pascual and Dunne [?].

Drossel and McKane proposed a model taking into account predator-prey dynamics (see [?, ?] and references within) implementing Arditi-Ginzburg Predator-Prey ratio-dependent efforts [?] and traits evolution (the proportion of a prey p in a predator P's is given as the score of a traits matching matrix). In [?] they implemented other predator-prey dynamics, e.g., Lotka-Volterra, and observed that the model food web outcome didn't fit the empirical observed data. The model has been developed in order to study immigration-speciation processes [?], and complexity-stability relation [?].

Caldarelli and Garlaschelli proposed a general framework for the ecological network evolution, the Webworld model [?]. Cannon et al. introduced a Lotka-Volterra niche model starting from a void universe with a single steady source of biomass [?] and gave an interpretation of two peculiar motifs in terms of network expansion within total biomass constraint. A food web model implementing niche model and the hypothesis that speciation and extinction rates decrease with increasing body mass was proposed in 2005 by Rossberg et al. [?] and then further developed [?] implementing the signal of an evolutionary history where new species avoid to compete with their ancestors.

It would be useful to reconstruct foodwebs structure along their historical evolution. Although detailed direct observations are yet to come, some attempts have been done, see Roopnarine et al. on paleocommunities [?, ?] or Dunne et al. for the Cambrian period [?]. Another possibility has been introduced by Doi et al. [?]: the author examined lakes spanning six order of magnitude different ages, and hence showing the result of shorter or longer evolution times.

Extinction events, hybrids and Phylogenetic Diversity   Extinction events prune the trees of life. This results in a loss of phylogenetic diversity, which depends clearly on which species goes extinct and which are retained.

A basic model which is widely used to predict on the future biodiversity

of a species community under extinction threat is the Field of Bullets model, introduced by Raup in 1984 [?]. Since its introduction the model has been variously generalised and applied with good results.

An immediately related problem asks, under limited resources, to find a conservation policy maximizing the expected biodiversity. A first answer comes from just considering the number of saved species. Prioritizing the length of the saved branches means taking into account Phylogenetic Diversity (Faith [?]): this approach was introduced by Witting and Loeschcke [?].

Weitzmann [?] considered probabilities of survival and budget limitation and introduced the so-called "Noah's Ark Problem". Hartmann and Steel [?] showed that, under certain limitation, we can optimally solve the problem through a greedy algorithm.

Bordewich and Semple [?] considered an important generalization of the Noah's Ark Problem where conservational policy can be applied to groups of species, or nature reserves, hence introducing relationship between the survival probability of different species.

Recently, Billionnet [?] showed that fast mixed-integer linear programming techniques can give a near-optimal solution to both the original Noah's Ark Problem and its nature reserves generalization. This techniques also yield an upper bound to the optimal solution.

With a different approach, a relationship between the survival probability of species taking into account their phylogenetic diversity (in terms of shared common evolutionary history) was introduced as a trait-based extinction model, introduced by Steel and Faller [?, ?]: the model allowed trait evolution along the tree branches following a markov chain model.

An explicit introduction of ecological constraints into a field of bullets extinction model has been proposed by Moulton et al. [?], where the authors used the concept of "viability", that is taking into account that every predator species has to predate at least on a species to survive. The model complexity has been studied by Faller et al. [?].

Species stability Foodwebs have been described in terms of ecosystem stability under a perturbation regime (which means, altering the population dynamic parameters, see for example Solé and Bascompte [?] for a recent

overview). An usual approach is to consider linear or approximately linear population dynamics – i.e., systems were the applied perturbation can be considered to produce a linear effect on the stable solution. As summarized by Solé and Bascompte, "Stability seems to be associated with diversity, but the exact nature of such association has been a matter of debates for decades" and so it is currently, we can argue.

Working on protein transcription networks and neuron networks, Prill et al. [?] gave a numerical characterization of 3-motifs and 4-motifs in terms of stability through a Structural Stability Score, defined as the probability that the dynamical system corresponding to a given motif relaxes monotonically to the stable equilibrium after a small linear perturbation. Significantly the authors found a positive correlation between the SSS score of a motif and its over or under representation in an empirical network.

Chapter 3

Research objectives and approach

We aim to discover the relationship between phylogenetic evolution and foodweb evolution.

A main focus of the thesis trying to disentangle the correlation between basic trophic network structural elements, such as elementary predator-preys linking, foodweb motifs, node or edge roles (as depicted in figure 3.1), and phylogenetic tree properties, such as phylogenetic diversity, pairwise distance between leaves, clades and clans. In this scenario, some foundational question are yet to be answered.
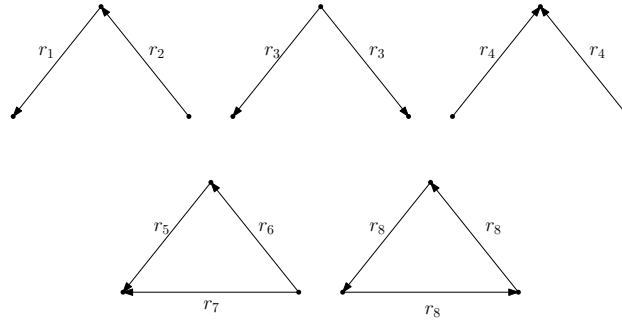
Figure 3.1: The different edge roles in the five most common predation motifs.

One problem, that will need to be solved, arises from the merging of data coming from two different empirical sources: genetic sequencing studies and foodwebs observation. The usual empirical data are given in the form of an ecological community which phylogeny is reconstructed and the species are mapped in a foodweb. This gives a straightforward mapping of the phyloge-

netic tree leaves into the trophic relationship graph. The food can be variably detailed in terms of guilds census (how many extant species and trophic relationship are present in the given foodweb) and taxonomic resolution (how coarse is the identification of the guilds taxonomy). Analog uncertainty can be present in the phylogeny reconstruction, which may vary from a topological tree given by the taxonomic classification to a tree reconstructed from the actual genome sequencing of the observed species.

3.1 Coupling ecological roles and phylogenetic structure

3.1.1 Foodweb and tree motifs

Let τ be the phylogenetic tree of the species in a certain ecological community and ω its trophic relationship graph, the foodweb. Call $l_1, l_2, \ldots, l_n$ the leaves of τ, we will use the same name to identify the species in the foodweb.
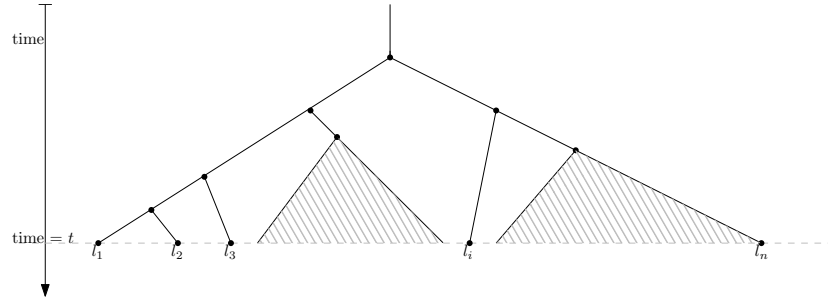
Figure 3.2: A Phylogenetic Tree, the shadowed zones hides other branches and leaves.

We will express ω in terms of its community adjacency matrix M_ω which elements m_ω^{ij} are equal to 1 if and only if there's a link between node i and j, and hence if and only if the species l_i predates on the species l_j. Through the tree τ we can define a phylogenetic distance matrix M_{PD} which elements m_{PD}^{ij} are given by the phylogenetic distance Pd_{ij} between the species l_i and l_j. As $Pd_{ij} = Pd_{ji}$, M_{PD} is a symmetric matrix.

The phenomenon we are looking at is apt to two complementary analysis

Figure 3.3: In this foodweb, and in all the proposal, we draw a link from a node i to a node j whenever species i predates on j.

approaches: 1) looking from the phylogeny to the foodweb structure, 2) looking from the foodweb structure to the phylogeny.
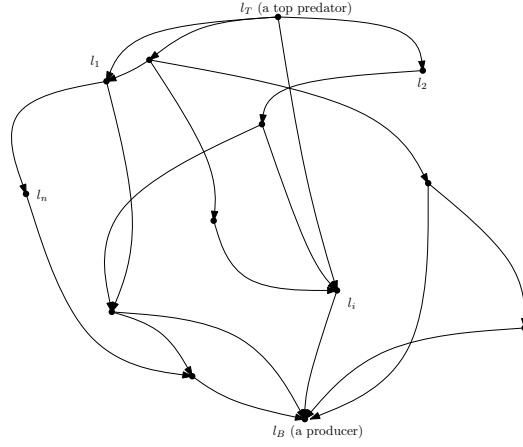
Couples   We will consider basic components of both the predation network and the graph, starting with couples and adding one element at time.

1. What is the probability of $m_{omega}^{ij} = 1$ given that the species l_i and l_j have phylogenetic distance $Pd_{ij} = k$? Or, similarly, give a good approximating function

$$f(k) \approx \mathbb{P}(m_{omega}^{ij} = 1 | Pd_{ij} = k).$$

Notice that, in a model where evolution and predation are independent, we would have

$$\mathbb{P}(m_{omega}^{ij} = 1 | Pd_{ij} = k) = \mathbb{P}(m_{omega}^{ij} = 1).$$

2. On the other hand, given $m_{omega}^{ij} = 1$, what is $\mathbb{E}\left[Pd_{ij}\right]$? In a null model we have $\mathbb{E}\left[Pd_{ij} | m_{omega}^{ij} = 1\right] = \mathbb{E}\left[Pd_{ij}\right]$.

Triplets   We will, then, study the identity mapping of foodweb elements triplets, called 3-motifs, into phylogenetic tree elements triplets, called triangles. We will call $\omega(i, j, \dots)$ the subgraph induced by the node $i, j, \dots$.

We will call *triangle of distances* a triplet $T_{ijk} = \{Pd_{ij}, Pd_{jk}, Pd_{ki}\}$ of mutual phylogenetic distances between three elements of the ecological community. Figure 3.4 represents a triangle between the species l_i, l_k, l_j.
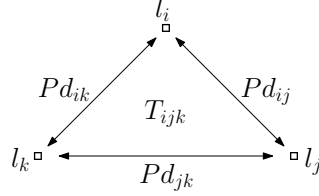
Figure 3.4: A triangle of phylogenetic distances.

In the basic five 3-motifs, the ones without mutual predation relationship, we have 8 topologically different relationship r_i, as depicted in figure 3.1. If a connected subgraph $\omega(i, j, k)$ is isomorphic to a certain motif with its edges in positions r_z, $r_{z'}$, $r_{z''}$ we will write

$$\omega(i, j, k) \to m(z, z', z'').$$

The question 2) has now a straightforward extension: we want to give

$$\mathbb{E}\left[T_{ijk} | \omega(i, j, k) \to m(z, z', z'')\right].$$

The generalization of question 1) requires more attention.

We should notice that each predation relationship, the link (i, j), can be part of various motifs including the species l_i, l_j and a third species completing the motif, as shown in figure 3.5.

Figure 3.5: The neighborhood of the link (i, j). In this example we will have $\underline{m}(i, j) = \{0, 1, 2, 1, 1, 1, 0, 0\}$.

We will call the *motif distribution* of a link (i, j) the vector $\underline{m}(i, j) = \{c_1, c_2, \ldots, c_8\}$ where c_i is given by the number of times the links (i, j) occurs

in position r_i. Each motif $\omega(i,j,k)$ is associated to a triangle of distances T_{ijk} and to each motif distribution $\underline{m}(i,j)$ is associated a vector of triangles $T_{\underline{m}(i,j)}$.
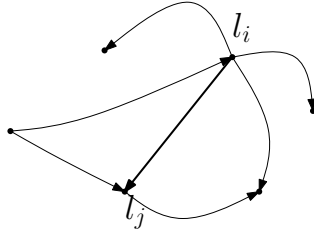
When considering the linking between two species in this scenario we should take into account also their 3-motifs neighborhood. Hence, the probability of having a predation relationship (i,k) may be influenced by the phylogenetic distance between l_i and l_k, the motif distribution such a relationship would have and all the triangles of distances induced by the motifs containing l_i and l_j if that relationship would exist.

Question 1) in this scenario asks: what's the probability of $m^{ij}_{omega} = 1$ given phylogenetic distance $Pd_{ij} = k$, motif distribution $\underline{m}(i,j) = \vec{v}$ and vector of triangles $T_{\underline{m}(i,j)}$. Or, similarly, give a good approximating function

$$f(k, \vec{v}, T_{\vec{v}}) \approx \mathbb{P}(m^{ij}_{omega} = 1 | Pd_{ij} = k, \underline{m}(i,j) = \vec{v}, T_{\underline{m}(i,j)} = T_{\vec{v}}).$$

Again, notice that, in a null model, the motif distribution should have no role.

3.1.2 Does predation separate clans?

Let's observe again a foodweb ω and the phylogenetic tree τ whose leaves set $L = \{l_1, l_2, \dots\}$ is a subset of the elements of ω. Explicitly denote the mapping of the leaves in ω as $p(l_i)$.

For each species l_i we can define one couple of subsets in the foodweb and one couple in the phylogenetic tree. Indeed, we can split the neighbourhood of $p(l_i)$ in ω, which is denoted $\star(p(l_i))$, in the set of species predating $p(l_i)$, denoted $T(p(l_i))$, and the set of species $p(l_i)$ predates, $B(p(l_i))$. Hence $\star(p(l_i)) = T(p(l_i)) \cup B(p(l_i))$. In a similar fashion, consider l_i and the subset of L given by the mapping of the species in $\star(p(l_i))$ into the phylogenetic tree, that is $p^{-1}(\star(p(l_i)))$, which we will denote as $\star(l_i)$. We define a rooted subtree by pruning all the species but the ones in $star(l_i)$ and placing the subtree root in l_i, which we denote $\tau_{|\star(l_i)}$. The subtree can be split in two principal clades, $C_r(l_i), C_l(l_i)$, such that $\tau_{|\star(l_i)} = C_r(l_i) \cup l_i \cup C_l(l_i)$.

We can, at this point, ask what is the distribution of the predators and preys of $p(l_i)$ on the two clades $C_r(l_i), C_l(l_i)$.

As a working null hypothesis, let's assume the predation relationship is a random process independent from the evolutionary history. Consider a species l_i, in the foodweb ω. Suppose l_i has in-degree $\delta^i(p(l_i))$ and out-degree $\delta^o(p(l_i))$. In- and out-degree of the species gives respectively the size of the l_i predators set $T(p(l_i))$ and of the l_i preys set $B(p(l_i))$. Predators and preys of l_i are distributed uniformly at random on the leaves of τ. This means that, taking a leaf from τ, the probability of getting a l_i predator, a l_i prey or a species that as not direct trophic relationship with l_i depends only the relative size of L, $B(p(l_i))$ and $T(p(l_i))$: drawing the lead uniformly at random or with a bias based on the tree structure doesn't change.

Calling $C_r(l_i)$ the larger of the two clans, the number of predators in it, which is the size of the set of leaves $C_r(l_i) \cap B(l_i)$, will be distributed as a hypergeometric random variable. A similar reasoning can be proposed for the prey.

A more refined null model can be retrieved by considering the probabilistic niche model [?] or the evolutionary niche model proposed by Ingram [?]: indeed, we can study the distribution of predators and preys on the two principle clades, assuming that, instead of being independent from the evolutionary history, the predation is shaped by a probabilistic niche model, where the niche position is a quantitative trait depending on the evolutionary history of the species.

Moreover, we can compare the model data with the data retrieved from the empirical observed network.

3.2 Toward a simple speciation/predation model

An insight into the mechanisms relating foodwebs evolution and speciation/extinction events can be acquired through the development of a suitable model which will be studied both from an analytical point of view and software implemented.

Consider a certain ecological community given by a set of species C_t at time t, let τ_t be its phylogenetic tree at time t and ω_t its trophic relationship graph (the foodweb) at time t. Call $l_1, l_2, \ldots, l_n$ the leaves of τ_t.

We will use the same naming also for the species in ω_t without fear of

confusion[1].

Evolution event   Each species l_i can be chosen for a speciation event at the same rate σ. When the species l_i is chosen at time t for speciation it gives birth to two species l_i' and l_i'', see fig. 3.6. Hence, we will observe the building of a classical Yule random tree.
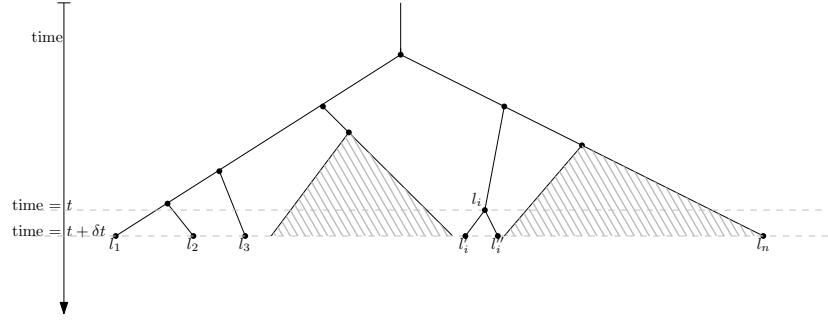
Figure 3.6: The phylogenetic tree τ at time $t + \delta t$ has two new species, l_i' and l_i''.

Imperfect duplication   To keep the model to the most simple case possible, we will assume that the place of one of the two siblings, l_i'', in $\omega_{t+\delta t}$ will duplicate exactly the place of l_i in ω_t. The other daughter will get rid of one of her ancestor predators with probability p and, independently from this, will add one new prey to her diet with probability q.

In this model the predator will be drawn uniformly at random in the set $P(l_i)$ of the predators of l_i and the prey will be chosen uniformly at random from $pp_t(l_i)$, which is the set of species l_p in C_t such that were no path from l_p to l_i. Hence, we do not allow for the formation of predation chain relationship. The mechanism is depicted in figure 3.7

What happens to the in- (and out-) degree of the graph during a speciation event?

If both siblings just duplicated their mother,

$$\delta_{l_i'}^{in}(t + \delta t) = \delta_{l_i''}^{in}(t + \delta t) = \delta_{l_i}^{in}(t) \ .$$

---

[1]In a more formal way we should introduce a bijective mapping i of the leaves l_i of τ_t to the species x_i in ω_t, so that $x_i = i(l_i)$.
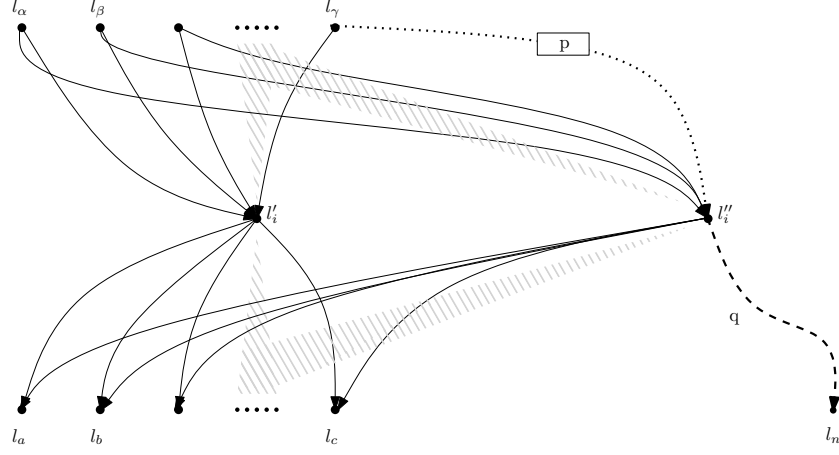
Figure 3.7: The rewiring of a species' foodweb neighbourhood during a speciation event in the imperfect duplication model.

At the same time all predators of l_i will acquire one more prey and all prey of l_i will have one more predator. Hence

$$\delta_{tot}^{in}(t + \delta t) = \delta_{tot}^{in}(t) + \delta_{l_i}(t).$$

If l_i' get rid of a predator, l_i' in-degree decreases by one and the out-degree of the lost predator decreases by one. On the other hand, if l_i' acquire a new prey, l_i' out-degree increases by one and the in-degree of the new prey increases by one. Hence, if l_i speciates at time t and we don't have other speciation before $t + \delta t$, we have that

$$\mathbb{E}\left[\delta_{tot}^{in}(t + \delta t)\right] = \delta_{tot}^{in}(t) + \delta_{l_i}(t) + q - p \tag{3.1}$$

which reduces, if $p = q$, to

$$\mathbb{E}\left[\delta_{tot}^{in}(t + \delta t)\right] = \delta_{tot}^{in}(t) + \delta_{l_i}(t) . \tag{3.2}$$

Moreover, naming $\widehat{\delta^{in}}(t)$ the mean degree at time t, during one speciation event occurring in the interval $[t, t + \delta t]$, assuming no other speciation event happen in that time, and remembering that $\hat{\delta}(t) = 2\widehat{\delta^{in}}(t)$ for each time t, we'll have that

$$\mathbb{E}\left[\widehat{\delta^{in}}(t + \delta t)\right] = \frac{(|C_t| + 2)\widehat{\delta^{in}}(t)}{|C_t| + 1} + \frac{q - p}{|C_t| + 1} \tag{3.3}$$

which reduces, for $p = q$, to

$$\mathbb{E}\left[\widehat{\delta^{in}}(t + \delta t)\right] = \widehat{\delta^{in}}(t) + \frac{\widehat{\delta^{in}}(t)}{|C_t| + 1} \tag{3.4}$$
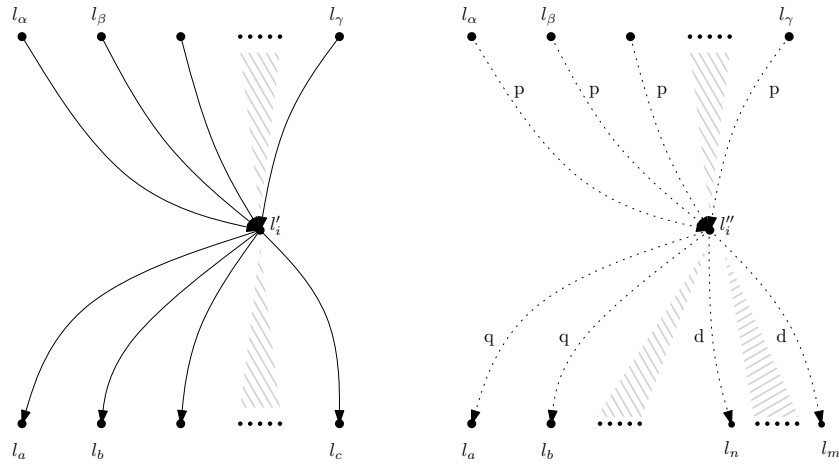
Figure 3.8: The rewiring of a species' foodweb neighbourhood during a speciation event in the niche partition model.

Random partition Another rewiring mechanism for the foodweb during a speciation event is the niche partition model. As before one of the two siblings keep the same place of her ancestor in the foodweb. The active daughter preserves each of her ancestor predator with a probability p (that is, gets rid of it with probability $1 - p$) and preserves each prey with probability q (that is, loses it with probability $1 - q$). Moreover, the daughter which changing its niche may add new prey in her diet drawing from the set of potential prey $pp_t(l_i)$ (which, similarly to what done in the previous model, is the set of species l_p in C_t such that aren't paths from l_p to l_i). The active daughter add each species in $pp_t(l_i)$ to her diet with probability

$$d = \frac{k}{|pp_t(l_i)|}$$

so that the overall probability of adding at least a new prey is

$$\mathbb{P}(\text{add at least one species}) = 1 - (1 - d)^{|pp_t(l_i)|} .$$

The mechanism is depicted in figure 3.8.

3.3 Extinction under complex ecological constraints

The most common simple mathematical model of extinction events are given by the *field of bullets* models and its generalization. A problem closely related is finding the, under a set of properly defined constraints, i.e. a cost/budget bound, the conservation policy maximizing the expected phylogenetic diversity.

A natural generalization of this approach may be given taking into account hybridization events and ecological constraint. Natural question regards the expected Phylogenetic Diversity, its variance and its comparison with the classical models.

Insights can be gathered from both a pure analytical approach, working with artificial data, and from the application of the models to empirical observed species community. Preliminary results shows that the behaviour of models where the extinction depends on ecological constraints and hybridization events are allowed is sensibly different from the classical models.

We just show briefly two possible initial model, which will be developed during the thesis research.

clade size weighted extinction

Very poor branches of the tree of life, such as just saved from extinction species or small isolated population may have lower probability to survive from a changing habitat.

We propose a model of random extinction which takes into account this issue.

Let τ be the phylogeny we are considering and x_i its leaves We will call C_r^0 and C_l^0 the two principal clades at time 0.

At each extinction event either a species or a clade goes extinct. To select which one to prune we think of extinction as an event running from the root to the leaves and stopping randomly. At each bifurcation the events steps on, it choose to stop with probability

$$\mathbb{P}(\text{stops at } C^y) = \left(\frac{1}{|C^y|}\right)^\alpha$$

or decides to proceed on the right or left clade with probability

$$\mathbb{P}(C^y_r) = \left(\frac{|C^y_r|}{|C^y_l|}\right)^\beta \mathbb{P}(C^y_l)$$

where C^y_r (C^y_l) is one of the two principal clades descending from C^y. A more precise notation can be given registering the r and l path followed in order to come to C^y.

When the extinction walk ends, all the species descending from the selected node go extinct.

In general, varying the values of β and α (as long as they're both positive), a continuous of scenario where isolated branches goes extinct can be modeled. Notice that if $\alpha \to \infty$ and $\beta = 0$ then the model is simply the classic Field of Bullets model.

Survival through hybridization

Motivation Let's consider a very simple case where we observe two groups of closely related taxa (A, B) and (B, C). Each of them have almost null probability of survival if they are not chosen for a conservation policy and probability p of survival otherwise. Moreover there's a probability $H(A, \neg B)$ that the clade (A, B) generate an hybrid $h_a b$ when A is protected but B is not; there's a probability $H(A, B)$ that the clade (A, B) generate an hybrid h_{ab} when both A and B are protected; there's a probability $H(\neg A, \neg B)$ that the clade (A, B) generate an hybrid h_{ab} when both A and B are not protected. We'll use a similar notation for the clade (C, D) or for hybrid coming from a couple of non-closely related species (as A and D or B and C, which are not considered in the Becker et al. paper).

Notice that, without hybridization, protecting the species A and C (scenario 1) brings to a greater expected biodiversity than saving species A and

B (scenario 2). On the other hand, if we allow for hybridization events, the previous inequality can be shown to be no more valid.

We will now analyze the expected Phylogenetic Diversity scenario, in the case with hybridization, in the spirit of Becker et al. A careful choice should be made when defining the weight of a hybrid species. We will introduce our choice with a suitable example.
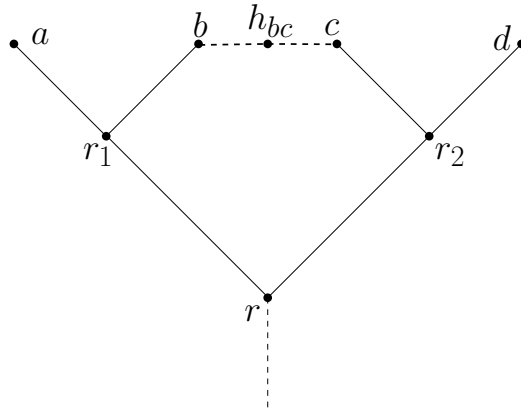
Figure 3.9: A phylogenetic tree where the species B and C generated an hybrid h_{bc}.

Shared history proportion   Let h_{bc} be a hybrid generated by the species B and C, as in figure 3.9. The hybrid will inherit a certain proportion of the genetic sequence of B, while the remaining part will be inherited from C. Let δ be the proportion of the genetic sequence of h_{bc} coming from B, we will write $h_{bc} = \delta C + (1 - \delta)B$, with $0 < \delta < 1$.

When computing the expected PD, we will consider that a part of the evolutionary history of C is conserved in h_{bc}, precisely a proportion equal to δ, and, at the same time a proportion $(1 - \delta$ of the evolutionary history of B is conserved in h_{bc}. These can be expressed imposing that the branch connecting h_{bc} to C is attached at a δ of its length, as in figure 3.10.

Given the length of the various branches in the original phylogeny (the one without the hybrid) and the survival probability for each of the extant species (and the hybrids), it is straightforward to compute the expected Phylogenetic Diversity of the tree.
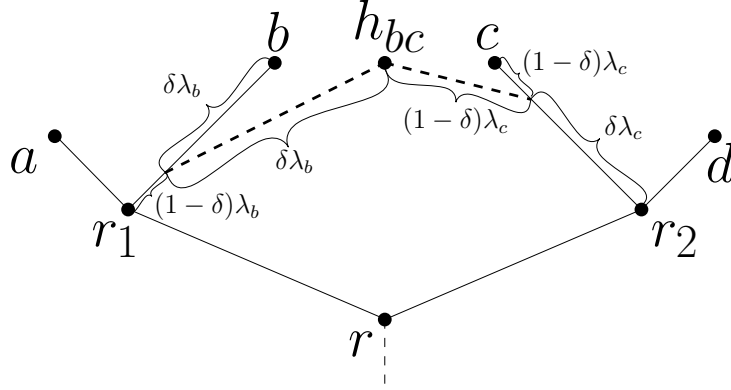
Figure 3.10: The hybrid h_{bc} conserve δ $((1-\delta))$ of the evolutionary history of species C (B), nearby the branches we show their length.

We will model hybridization through a random process occurring with a probability depending on the evolutionary history of the species

$$\mathbb{P}(H(i,j)) = \frac{1}{\text{k } d(i,j)}$$

where k is a scale factor ($\text{k} \geq 1$), $d(i,j)$ is the phylogenetic distance between species i and j normalized so that the shortest branch in the phylogeny has length 1, so to have $\min_{i,j}(d(i,j)) \geq 2$. Moreover, if a species is extinct or doomed, that is, has zero survival probability, its probability to generate a hybrid is also zero.

In the case where a hybrid may be originated, its PD contribution will be weighted by both the hybridization probability, $H(i,j)$, and the survival probability, p_h, of the hybrid.

Appendix A

Work Plan

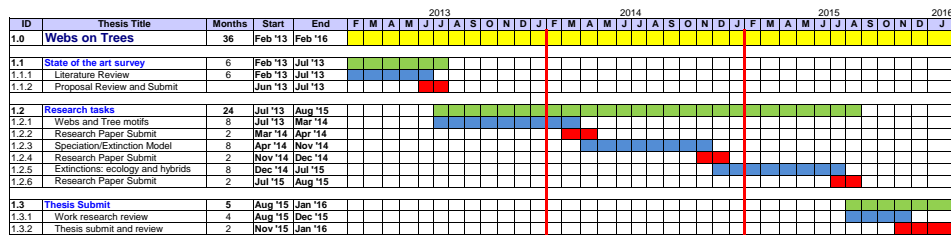| ID	Thesis Title	Months	Start	End	2013												2014												2015												2016
					F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J
1.0	**Webs on Trees**	36	Feb '13	Feb '16																																				
1.1	State of the art survey	6	Feb '13	Jul '13																																					
1.1.1	Literature Review	6	Feb '13	Jul '13																																					
1.1.2	Proposal Review and Submit		Jun '13	Jul '13																																					
1.2	Research tasks	24	Jul '13	Aug '15																																					
1.2.1	Webs and Tree motifs	8	Jul '13	Mar '14																																					
1.2.2	Research Paper Submit	2	Mar '14	Apr '14																																					
1.2.3	Speciation/Extinction Model	8	Apr '14	Nov '14																																					
1.2.4	Research Paper Submit	2	Nov '14	Dec '14																																					
1.2.5	Extinctions: ecology and hybrids	8	Dec '14	Jul '15																																					
1.2.6	Research Paper Submit	2	Jul '15	Aug '15																																					
1.3	Thesis Submit	5	Aug '15	Jan '16																																					
1.3.1	Work research review	4	Aug '15	Dec '15																																					
1.3.2	Thesis submit and review	2	Nov '15	Jan '16																																					

Figure A.1: Gantt diagram of the PhD work tasks. See last page for better resolution

ID	Thesis Title	Months	Start	End
1.0	Webs on Trees	36	Feb '13	Feb '16
1.1	State of the art survey	6	Feb '13	Jul '13
1.1.1	Literature Review	6	Feb '13	Jul '13
1.1.2	Proposal Review and Submit		Jun '13	Jul '13
1.2	Research tasks	24	Jul '13	Aug '15
1.2.1	Webs and Tree motifs	8	Jul '13	Mar '14
1.2.2	Research Paper Submit	2	Mar '14	Apr '14
1.2.3	Speciation/Extinction Model	8	Apr '14	Nov '14
1.2.4	Research Paper Submit	2	Nov '14	Dec '14
1.2.5	Extinctions: ecology and hybrids	8	Dec '14	Jul '15
1.2.6	Research Paper Submit	2	Jul '15	Aug '15
1.3	Thesis Submit	5	Aug '15	Jan '16
1.3.1	Work research review	4	Aug '15	Dec '15
1.3.2	Thesis submit and review	2	Nov '15	Jan '16

Timeline headers: 2013, 2014, 2015, 2016 with month columns (F M A M J J A S O N D J ...)