Landscapes on Spaces of Trees

Oliver Bastert [a], Dan Rockmore [b], Peter F. Stadler [c,d], and Gottfried Tinhofer [a]

[a] *Zentrum Mathematik, TU München*
D-80290 Muenchen, Germany
{bastert,gottin}@mathematik.tu-muenchen.de

[b] *Department of Computer Science, Dartmouth College*
6211 Sudikoff Laboratory, Hanover, NH 03755-3510, USA
rockmore@cs.dartmouth.edu

[c] *Institut für Theoretische Chemie, Universität Wien*
Währingerstraße 17, A-1090 Wien, Austria
*Phone: **43-1-4277-52737, Fax: **43-1-4277-52793,*
E-Mail: studla@tbi.univie.ac.at

[d] *The Santa Fe Institute*
1399 Hyde Park Road, Santa Fe, NM 87501, USA

Abstract

Combinatorial optimization problems defined on sets of phylogenetic trees are an important issue in computational biology, for instance the problem of reconstruction a phylogeny using maximum likelihood or parsimony approaches. The collection of possible phylogenetic trees is arranged as a so-called Robinson graph by means of the nearest neighborhood interchange move. The coherent algebra and spectra of Robinson graphs are discussed in some detail as their knowledge is important for an understanding of the landscape structure. We consider simple model landscapes as well as landscapes arising from the maximum parsimony problem, focusing on two complementary measures of ruggedness: the amplitude spectrum arising from projecting the cost functions onto the eigenspaces of the underlying graph and the topology of local minima and their connecting saddle points.

Key words: Fitness landscapes, phylogenetic trees, spectral graph theory, parsimony problem.

Subject Classification: 05C38, 05C85.

Manuscript *2 January 2001*

1 Introduction

Trees are used extensively in many fields to depict hierarchical relationships. In biology evolutionary relationships between species or individual genes are customarily represented in this way. The building of *phylogenetic trees* from sequence (and, more recently, structural) data is hence a central problem in computational biology. The vertices of a phylogenetic tree represent taxonomic units, the graph's topology delineates the genealogical relationships between them, and the branch lengths reflect the time of divergence. Many methods exist for the construction of phylogenetic trees. The more sophisticated among them seek those trees in which the taxonomic units evolve with the least evolutionary change [14] (most parsimonious trees) or trees with the maximum likelihood given a stochastic model of sequence evolution [16].

In mathematical terms we can rephrase the phylogeny reconstruction problem as follows: Let $\mathcal{A}$ be a set of extant taxonomic units or "species". A *phylogenetic tree on* $\mathcal{A}$ is an unrooted unordered tree all inner vertices of which have degree 3 and such that the leaves are uniquely labeled by the elements of $\mathcal{A}$. Furthermore, we are given a cost function f that allows us to determine "how well" a particular tree fits the genealogical relationships among the species in $\mathcal{A}$. A particular example, the so-called *parsimony score* will be discussed in some detail in section 4.2. The search for the optimal tree is hence recast as a combinatorial optimization problem: Given given $\mathcal{A}$, $\mathfrak{B}_{\mathcal{A}}$, and the fitness function f on $\mathfrak{B}_{\mathcal{A}}$, finding the optimal phylogenetic tree consists of minimizing f. We write $\mathfrak{B}_n$ for $\mathfrak{B}_{\{1,2,\dots,n\}}$.

The basic variants of these tree reconstruction problems are all NP-complete [21, 12]. There are

$$|\mathfrak{B}_n| = \prod_{k=1}^{n-2}(2k-1) = 1 \cdot 3 \cdot 5 \cdot (2n-5) \tag{1}$$

phylogenetic trees on n species [41, 39]. Exhaustive search is hence limited to small values of n, say $n \leq 12$ with $|\mathfrak{B}_{12}| \approx 6.5\,10^8$ different trees. Computer programs such as Felsenstein's `PHYLIP` hence resort to heuristic searches [17].

We will be concerned here with qualitative features of landscapes on spaces of phylogenetic trees. From the mathematical point of view, a *landscape* consists of a finite set X of *configurations*, a *fitness functions* $f : V \to \mathbb{R}$, and a neighborhood relation $\mathcal{N}$ on X. In other words, $(X, \mathcal{N})$ forms a (finite) graph Γ, and the landscape is viewed as function of its vertices [42].

The focus of this paper is to cast the reconstruction problem for phylogenetic trees in an algebraic framework. Early work of Robinson [39] is used to construct a graph on $\mathfrak{B}_n$ by using the *nearest neighbor interchange* move

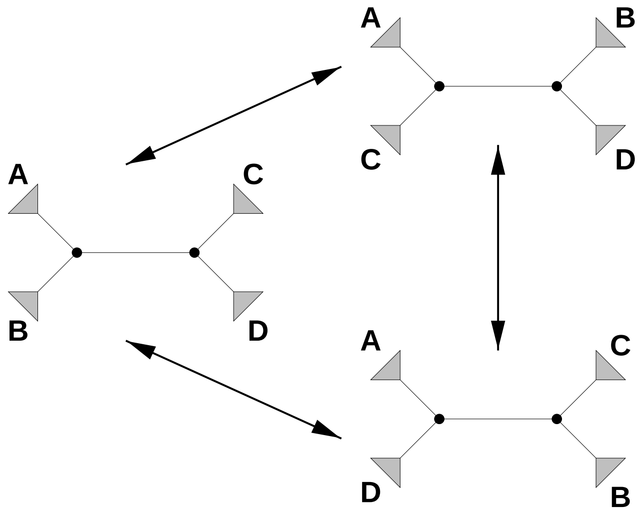

Fig. 1. Nearest neighbor interchange moves

set. In this setting, cost functions can be studied as elements in an associated *coherent algebra* and associated fitness landscapes can be analyzed according to their *amplitude spectrum* [25, 42, 29] and by means of their *barrier trees* [22, 20]. The maximum parsimony problem is used as an illustrative example.

2 Robinson Graphs

2.1 Definition

If $T \in \mathfrak{B}_n$, we write $E(T)$ for the edge set of the tree and $V(T)$ for its vertex set. A vertex is *interior* if it is not a leaf. An edge is interior if its connects two interior vertices. We denote the set of interior vertices and edges by $\hat{V}(T)$ and $\hat{E}(T)$, respectively. For later reference we note that $|V(T)| = 2n-2$, $|E(T)| = 2n-3$, $|\hat{V}(T)| = n-2$, and $|\hat{E}(T)| = n-3$. The interior subtree $T^\circ = (\hat{V}, \hat{E})$ of T consists of the interior vertices and edges.

A *nearest neighbor interchange* (nni) at the interior edge $e = \{x, y\}$ consists of exchanging a subtree attached to x with a subtree attached to y, see Fig. 1. Clearly only two of the four possible swaps yield distinct trees. Since there are $|\hat{E}(T)| = n-3$ interior edges for each $T \in \mathfrak{B}_n$, there are $2(n-3)$ possible nni moves, each leading to a different tree.

Clearly, an nni can be undone by a second nni using the same edge. Thus we may regard $\mathfrak{B}_n$ as the vertex set of an undirected graph Γ_n in which edges connect trees that differ by a single nni move. The following basic properties of Γ_n are verified already in Robinson's original paper [39]:

(i) Γ_n is $2(n-3)$ regular.
(ii) Γ_n is connected.
(iii) Γ_n is *not* distance degree regular for $n \geq 6$ because the number $w_2(T)$ of neighbors in distance 2 depends on the topology of the interior interior subtree T° of $T \in \mathfrak{B}_n$. $\Gamma_4 = K_3$ and Γ_5 are distance degree regular.

The diameter of Γ_n is considered in detail in [34].

Note that nni is by no means the only reasonable edit operation on phylogenetic trees. Other commonly used move sets are "Subtree Pruning and Regrafting" spr and "Tree Bisection-Reconnection" tbr, see [45]. It can be shown that a nni-move is also a spr-move and a spr-move is also a tbr-move. The Robinson graphs are hence subgraphs of the spr and tbr graphs, which will be studied elsewhere.

2.2 Coherent Algebras

A matrix algebra is a linear space of matrices which is closed with respect to matrix multiplication. The full matrix algebra with rows and columns indexed by elements of X is denoted Mat_X. Various subalgebras appear as subsets of Mat_X specified by requiring certain constraints on the entries.

Let $\mathbf{I}$ be the unit matrix of Mat_X and $\mathbf{J}$ the matrix with all entries equal to 1. For two matrices $\mathbf{A}$ and $\mathbf{A}'$ define the *component-wise product* $\mathbf{A} \circ \mathbf{A}'$ by

$$(\mathbf{A} \circ \mathbf{A}')_{ij} = A_{ij} A'_{ij}. \tag{2}$$

A sub-algebra $\mathcal{A}$ of Mat_X is called a *coherent* or a *cellular algebra* (on X) if it contains $\mathbf{I}$ and $\mathbf{J}$, and if it is closed with respect to the component-wise product and with respect to transpose. Coherent algebras have been introduced and studied first in [50, 49], and independently in [26, 27, 28]. Since then a rich theory has been built up around them in the literature. Today, the notion of coherent algebra and the equivalent notion *coherent configuration* (see [27]) are among the main tools of algebraic combinatorics. A friendly introduction to coherent algebras taking into account the interests of chemists is given in [31], while the paper [13] is written for mathematicians and covers the most important theoretical aspects. We list here some properties of coherent algebras, relevant for the analysis which follows. For proofs see [26].

(i) Every coherent algebra possesses a unique linear basis $\mathbf{A}_1, \ldots, \mathbf{A}_s$ consisting of $0,1$-matrices $\mathbf{A}_i$, called *the standard basis*, such that

$$(a) \quad \mathbf{I} = \sum_{i=1}^{t} \mathbf{A}_i \text{ for some } t,\ 1 \le t < s, \text{ and} \tag{3}$$

$$(b) \quad \mathbf{J} = \sum_{i=1}^{s} \mathbf{A}_i \tag{4}$$

The second condition says that $\mathbf{A}_i \circ \mathbf{A}_j = 0$ for $i \neq j$.

(ii) For every basis matrix $\mathbf{A}_i$ there is a basis matrix $\mathbf{A}_{i'}$ such that $\mathbf{A}_{i'} = \mathbf{A}_i^{\mathsf{T}}$. Note that in general $i' \neq i$, the matrices $\mathbf{A}_i$ with $i > t$ are not necessarily symmetric.

(iii) The product of two basis matrices is a linear form

$$\mathbf{A}_i \mathbf{A}_j = \sum_{k=1}^{r} p_{ij}^k \mathbf{A}_k \tag{5}$$

with integral coefficients p_{ij}^k which are called the *structure constants* of the algebra.

(iv) According to (i.a) the sets $C_i = \{v \in X \,:\, (\mathbf{A}_i)_{vv} = 1\}$, $1 \le i \le t$, form a partition $\mathcal{C}_X$ of X, called the *cell partition*. The sets C_i are called the *cells* of the coherent algebra. The cell partition is *equitable*[1] with respect to (the graph associated with) every basis matrix $\mathbf{A}_i$.

(v) The basis matrices $\mathbf{A}_i$ can be considered as adjacency matrices of (in general directed) graphs $G_i = (X, E_i)$, the *basis graphs*. Their arc sets are $E_i = \{(u,v) \,:\, (\mathbf{A}_i)_{uv} = 1\}$. The sets E_i, $1 \le i \le s$, are called the *basis sets* of the algebra. They form a partition of $X \times X$ which is called a *coherent configuration*.

(vi) For each basis set E_i there are cells $C_j, C_k \in \mathcal{C}_X$ such that $E_i \subseteq C_j \times C_k$. Further, in this case,

$$\begin{aligned} |\{v \,:\, (u,v) \in E_i\}| &= \begin{cases} |E_i| \big/ |C_j| & \text{for } u \in C_j \\ 0 & \text{otherwise.} \end{cases} \\ |\{v \,:\, (v,u) \in E_i\}| &= \begin{cases} |E_i| \big/ |C_k| & \text{for } u \in C_k \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{6}$$

(vii) A coherent algebra for which $t = 1$ (or with other words, in which $\mathbf{I}$ is a basis matrix) is called *homogeneous*. In a homogeneous coherent algebra $C_1 = X$, so the cell partition is trivial.

The smallest coherent algebra containing the adjacency matrix $\mathbf{A}$ of a graph Γ is called the coherent algebra *generated* by Γ and denoted by $\langle\langle \Gamma \rangle\rangle$.

[1] A partition $\{C_1, \ldots, C_m\}$ of the vertex set of a graph is *equitable* if, for any choice of i and j, all vertices in C_i have the same number $\hat{\mathbf{A}}_{ij}$ of neighbours in C_j. The matrix $\hat{\mathbf{A}}$ is the *collapsed adjacencency matrix*.

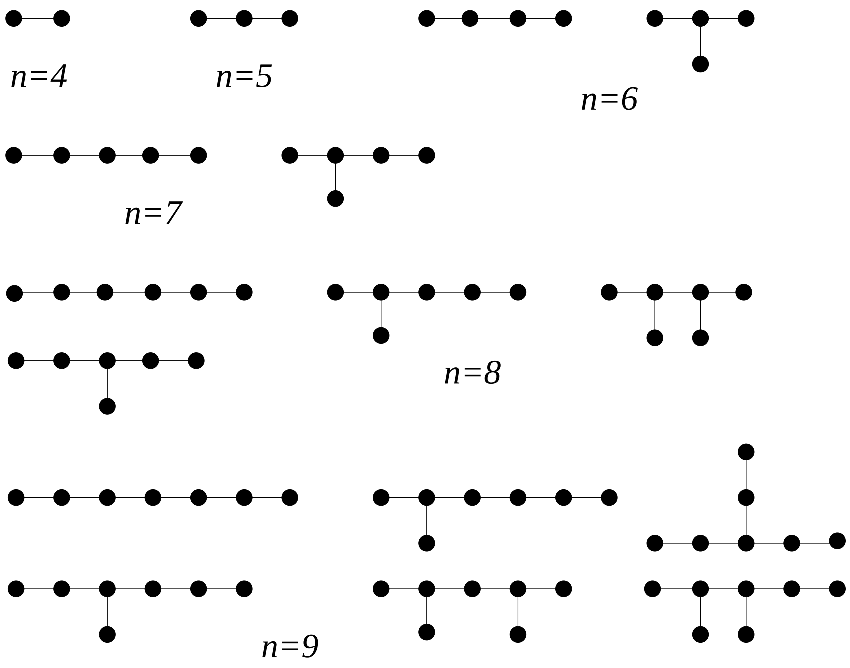

Fig. 2. Interior tree topologies for $n = 4$ to $n = 9$.

2.3 The Cell Partition of Robinson Graphs

The structure of the coherent algebra $\langle\!\langle \Gamma_n \rangle\!\rangle$ and its spectral properties can be understood in terms of natural symmetries of Γ_n.

Theorem 1 *Let T_1 and T_2 be two phylogenetic trees such that their interior trees are equal, $T_1^\circ = T_2^\circ$. Then there is an automorphism $\alpha : \mathfrak{B}_n \to \mathfrak{B}_n$ of Γ_n such that $\alpha(T_2) = T_1$.*

Proof. Let π be a permutation of $\{1, 2, \ldots, n\}$. Then the action of π on $\{1, 2, \ldots, n\}$ extends to an action α_π on $\mathfrak{B}_n$ by permuting the labels on the leaves. The action of α_π clearly commutes with any `nni` move, hence α_π is an automorphism of Γ_n. For each pair T_1, T_2 of trees with the same interior tree T° there is a permutation π such that $T_1 = \pi(T_2)$, hence trees with the same interior tree are contained in the same orbit of $\mathsf{Aut}[\Gamma_n]$. □

Definition 1 *The* interior tree partition *of $\mathfrak{B}_n$ consists of the classes of trees that have the same interior tree.*

Using definition 1 we can rephrase Theorem 1 as follows: The interior tree partition of $\mathfrak{B}_n$ is finer than the cell partition of the coherent algebra $\langle\!\langle \Gamma_n \rangle\!\rangle$.

For $n = 5$ there is only a single interior tree, which is isomorphic to the path P_3. By theorem 1 $\mathsf{Aut}[\Gamma_5]$ has only a single orbit, i.e., Γ_5 is vertex transitive. This is not true for $n \geq 6$. In Figure 2 we show the interior trees for $n = 4$ to $n = 9$.

A most appealing result on Robinson graphs is the following converse of Theorem 1:

Theorem 2 *The interior tree partition of* $\mathfrak{B}_n$ *equals the cell partition of the coherent algebra* $\langle\langle\Gamma_n\rangle\rangle$.

Proof. The proof of this theorem requires the knowledge of quite a few details of the structure of Γ_n. Therefore, we only give a short sketch here. For the complete proof we refer to [4].

Let $\mathcal{C}_n$ be the cell partition of Γ_n. As mentioned in 2.2(iv), $\mathcal{C}_n$ is equitable with respect to every basis matrix $\mathbf{A}_i$. This implies that the entries of any two rows (columns) of an arbitrary matrix $\mathbf{A}'$ in $\langle\langle\Gamma_n\rangle\rangle$ belonging to trees T and T' in the same cell of $\mathcal{C}_n$ must sum to the same amount. In particular, this means that for any k, the number of trees at distance k must be the same for both T or T'. We use this criterion several times in order to distinguish trees which belong to different cells.

Interior trees have vertices of degree 1, 2, or 3. Let their numbers be n_1, n_2, and n_3, respectively. These numbers are not independent. We have $n_2 = n_1 - 2$ and $n_3 = n - 2n_1$. In [39] it was shown that the number of trees at distance 2 from a given tree T depends only on the number n_1. From this we see immediately that n_1, n_2, and n_3 are constant on each cell of $\mathcal{C}_n$. Since a path is the only tree with $n_3 = 0$, all trees in Γ_n the interior tree of which is a path belong to a single cell, say to C_1. Since any two paths of the same length are isomorphic, C_1 is an orbit of $\mathsf{Aut}[\Gamma_n]$.

Any single nni can change the diameter of a tree by at most one unit. The diameter increases if and only if the nni is made using an edge having exactly one of its vertices on a longest path in the interior tree, while it decreases if and only if the nni is made using an edge which is part of all longest interior paths. This information together with the fact that the maximum diameter is realized in trees of C_1 only, can be used to prove (by downward induction) that the diameter is constant on each cell.

The remaining steps in the proof for the theorem are more involved. Essentially, it depends upon showing that the interior tree T° of a tree T is uniquely determined up to isomorphism by the number of its neighbors having larger diameter and by the isomorphism type of the interior trees of these neighbors. All details are given in [4]. $\square$

2.4 Small Robinson Graphs

For small values of n a complete analysis of the Robinson graphs is feasible.

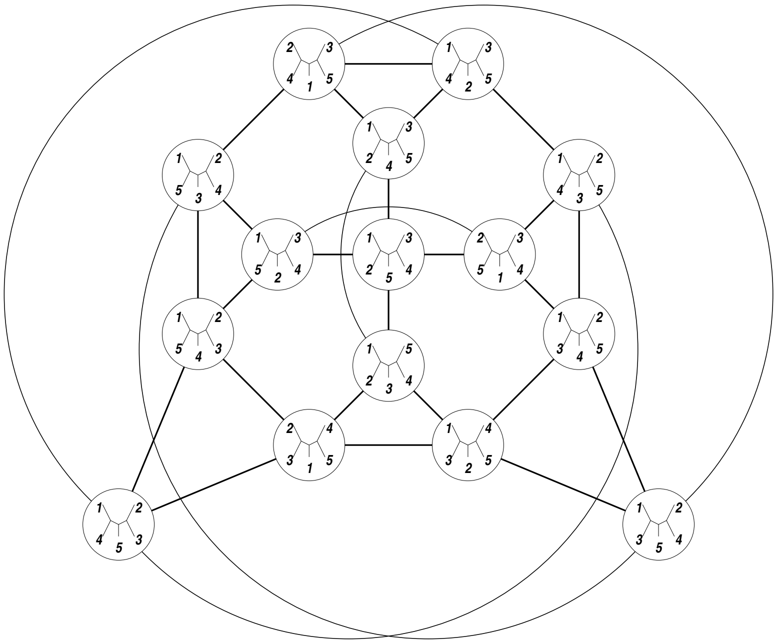

Fig. 3. $\Gamma_5 = L[P]$.

- $\Gamma_1 = \Gamma_2 = \Gamma_3 = K_1$ since there is only a single phylogenetic tree with $n = 1$, 2, or 3 end-vertices.
- $\Gamma_4 = K_3$, since there are only the three trees shown in Figure 1.
- Γ_5 is easily constructed explicitly, see Figure 3. It is distance regular with diameter 3. The collapsed adjacency matrix (structure matrix)

$$\hat{\mathbf{A}} = \begin{pmatrix} 0 & 4 & 0 & 0 \\ 1 & 1 & 2 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 4 & 1 \end{pmatrix} \tag{7}$$

(or, more precisely the off-diagonal entries known as the intersection array) identify Γ_5 as a known distance-transitive graph, namely the *line graph of the Petersen graph* $L[P]$ [6, Thm.7.5.3]. Since it contains K_5 as a minor it is not planar. The Laplacian spectrum of Γ_5 is easily computed explicitly:

j	$=$	0	1	2	3
Λ_j	$=$	0	2	5	6
$\mathrm{mult}(\Lambda_j)$	$=$	1	5	4	5

(8)

For $n \leq 9$ we were able to generate the Robinson graphs on the computer. For $n = 6, 7$ the adjacency matrix is still small enough so that the spectrum of the Laplacian can computed directly via matrix diagonalization. Additional

Table 1
Some numerical data on small Robinson graphs. The values for diamΓ_n are taken from [34].

n	$\|V\|$	deg	diam	Cells	$\dim\langle\langle\Gamma_n\rangle\rangle$	$\|\text{spec}\|$	$\Lambda_{\max}$
3	1	0	0	1	1	1	0
4	3	2	1	1	2	2	3
5	15	4	3	1	4	4	6
6	105	6	5	2	31	10	9
7	945	8	7	2	243	46	12
8	10395	10	10	4	?	213	15
9	135135	12	12	6	?	$\geq$1041	18

algebraic information can be obtained by computing the coherent algebra $\langle\langle\Gamma\rangle\rangle$ using Oliver Bastert's program `qweil` [2]. In particular, the algebra $\langle\langle\Gamma\rangle\rangle$ is not homogeneous (and hence not commutative [26]) for $n \geq 6$. Summary information on the coherent algebras of Γ_6 and Γ_7 can be found in Table 1.

Theorem 1 together with the following result from [44] sets the stage for an efficient way of computing the spectrum of Γ_n for a few larger values.

Theorem 3 *Let $\pi^{(\kappa)} = \{X_1^{(\kappa)}, \ldots, X_{s_\kappa}^{(\kappa)}\}$, $1 \leq \kappa \leq \nu$, be a sequence of equitable partitions satisfying the condition that for every $x \in X$ there is a κ with $\{x\} \in \pi^{(\kappa)}$. Then*

$$\bigcup_{\kappa=1}^{\nu} \text{spec}(\mathbf{C}^{(\kappa)}) = \bigcup_{\kappa=1}^{\nu} \text{spec}(\mathbf{R}^{(\kappa)}) \subseteq \text{spec}(\mathbf{A}) \subseteq \bigcup_{\kappa=1}^{\nu} \text{spec}(\mathbf{R}^{(\kappa)}), \tag{9}$$

The structure matrices $\mathbf{R}^{(\kappa)}$ and $\mathbf{C}^{(\kappa)}$ of $\pi^{(\kappa)}$ are defined by

$$\begin{aligned} \mathbf{R}_{ik}^{(\kappa)} &= \sum_{x \in X_k^{(\kappa)}} \mathbf{A}_{ux} \text{ independent of } u \in X_i^{(\kappa)} \\ \mathbf{C}_{ik}^{(\kappa)} &= \sum_{x \in X_i^{(\kappa)}} \mathbf{A}_{xy} \text{ independent of } y \in X_k^{(\kappa)} \end{aligned} \tag{10}$$

The equitable partitions obtained from different representatives T of the same class of the cell partition of $\mathbf{A}$ are equivalent in the sense that they yield the same structure matrices (possibly up to permutations of the indices). Thus we obtain the complete spectrum of $\mathbf{A}$ as the union of the spectra of the structure matrices $\mathbf{R}^{(\kappa)}$ where κ now indexes the classes of the cell partition of $\langle\langle\Gamma_n\rangle\rangle$. This is equivalent to computing equitable partitions for a single reference tree from each class of the interior degree partition.

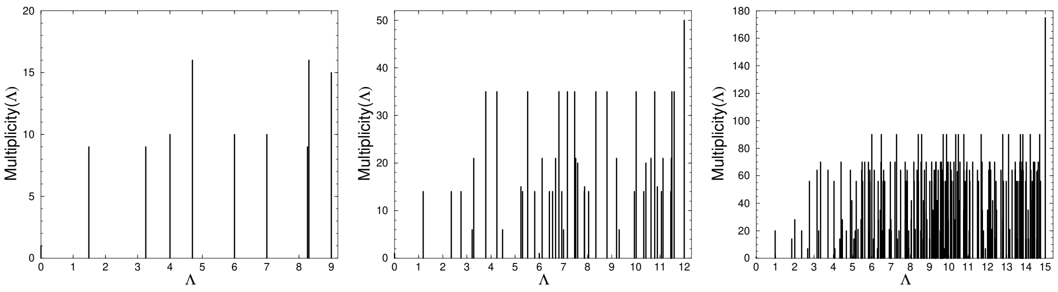

Fig. 4. Laplacian Spectra of the Robinson graphs Γ_6, Γ_7 and Γ_8

The point is that the equitable partition "anchored" at a tree T can be computed very efficiently in $\mathcal{O}(|E|\log|V|)$ time and $\mathcal{O}(|E|)$ space using `qstab` [3]. This has to be done only for a single representative of each class of interior trees. Furthermore, the dimension of the structure matrices $\mathbf{R}^{(\kappa)}$ is in general much smaller than $|V|$. The spectra of $\mathbf{R}^{(\kappa)}$ can either be computed directly, or one can try to find equitable partitions of $\mathbf{R}^{(\kappa)}$ and to compose $\mathrm{spec}(\mathbf{R}^{(\kappa)})$ from a collection of even smaller matrices, following the procedure outlined in [44]. The eigenvectors can then be obtained by numerically solving $(\mathbf{A}-\lambda\mathbf{I})x = 0$. The use of symmetry-based techniques for block diagonalization of adjacency matrices also has been used to analyze the spectra of Cayley graphs (see eg. [33]).

The results for $n = 6$, $n = 7$, and $n = 8$ are summarized in Figure 4. For $n = 9$ we obtain 6 inequivalent equitable partitions with 1534, 3610, 5901, 10815, 19698, and 21252 cells, respectively. The structure matrices of the smallest four partitions were diagonalized directly, yielding a lower bound of 1041 distinct eigenvalues.

3 Rugged Landscapes

3.1 Correlation Measures and Amplitude Spectra

All measures discussed in this section are invariant under affine transformation. Hence we may assume

$$\begin{aligned} \overline{f} &= \frac{1}{|X|}\sum_{x\in X} f(x) = 0 \\ \mathrm{Var}[f] &= \sum_{x\in X}\left(f^2(x) - \overline{f}^2\right) = \langle f, f\rangle = 1 \end{aligned} \tag{11}$$

Let $\Gamma = (X, \mathcal{N})$ be a graph with adjacency matrix $\mathbf{A}$. Then the diagonal matrix of vertex degrees $\mathbf{D}$ has diagonal entries $\mathbf{D}_{xx} = \sum_y \mathbf{A}_{xy}$. The canonical

Markov transition operator on Γ is $\mathbf{T} = \mathbf{A}\mathbf{D}^{-1}$. The "random-walk" autocorrelation function [48] of the landscape f satisfying (11) is [42]

$$r(s) = \langle f, \mathbf{T}^s f\rangle \tag{12}$$

From (12) one obtains the *correlation length* of f on Γ as $\ell = \sum_{s=0}^{\infty} r(s)$.

The *Laplacian matrix* of Γ is defined as

$$-\mathbf{\Delta} = \mathbf{D} - \mathbf{A} \tag{13}$$

is a particularly useful algebraic representation of Γ. It is a symmetric non-negative definite matrix with smallest eigenvalue 0, the multiplicity of which equals the number of connected components. For surveys on graph Laplacians see e.g. [36, 37] and the book [10].

In order to show $\mathbf{\Delta}, \mathbf{T} \in \langle\!\langle\Gamma\rangle\!\rangle$ it suffices to verify that $\mathbf{D} \in \langle\!\langle\Gamma\rangle\!\rangle$, which follows from the fact that the degree partition is coarser than the cell partition for any graph, see e.g. [31]. If Γ is regular, as is the case with the Robinson graphs, then $\mathbf{A}$, $\mathbf{T}$ and $-\mathbf{\Delta}$ have the same eigenvectors since $\mathbf{D}$ is a multiple of the identity matrix. Let $\{\varphi_k\}$ be an orthonormal basis of $-\mathbf{\Delta}$ with associated eigenvalues Λ_k. The decomposition

$$f = \sum_k a_k \varphi_k \tag{14}$$

is sometimes called a Fourier decomposition of the landscape. The *amplitude* $B(\Lambda)$ of f on the eigenspace $\mathcal{E}_\Lambda = \{\xi | -\mathbf{\Delta}\xi = \Lambda\xi\}$ is the *fraction* of the landscape variance that is contributed by the projection of f onto $\mathcal{E}_\Lambda$. Thus $B(\Lambda)$ is invariant under affine transformations $f \to \sigma f + \overline{f}$. Using (11) we can express $B(\Lambda)$ in terms of the the Fourier decomposition (14) in the form

$$B(\Lambda) = \sum_{k: -\mathbf{\Delta}\varphi_k = \Lambda\varphi_k} a_k^2 \tag{15}$$

for all eigenvalues $\Lambda \neq 0$. By (11) we have $\sum_k a_k^2 = 1$ and $B(0) = a_0^2 = \overline{f}^2 = 0$. The corresponding expressions for non-normalized landscapes can be found in [42]. In general we have $B(\Lambda) \geq 0$ and $\sum_\Lambda B(\Lambda) = 1$, with the lanscape becoming more rugged as the high-Λ amplitudes increase.

In the case of D-regular graphs, which of course includes the Robinson graphs, there are simple expressions for the correlation function $r(s)$ and the correlation length ℓ in terms of the amplitude spectrum [42]:

$$r(s) = \sum_\Lambda B(\Lambda)\left(1 - \frac{\Lambda}{D}\right)^s \qquad \text{and} \qquad \ell = D\sum_\Lambda \frac{B(\Lambda)}{\Lambda} \tag{16}$$

The amplitude spectrum is therefore a detailed correlation measure, with the landscape becoming more rugged the larger the high-Λ amplitudes are. A landscape that is – up to an additive constant – an eigenfunction of $-\mathbf{\Delta}$, i.e., for which $B(\Lambda) = 1$ for one eigenvalue $\Lambda \neq 0$ and $B(\Lambda') = 0$ for all $\Lambda' \neq \Lambda$ is called *elementary*. This notion is important because on the one hand many of the best studied combinatorial optimization problems form elementary landscapes on their "natural" configurations graphs [24, 42, 43], and on the other hand eigenfunctions of the graphs Laplacian have a number of distinct geometric properties: All their local minima have a value below the landscape average [24], and they satisfy a version of Courant's nodal domain theorem, implying that ruggedness indeed increases with the location in the spectrum of $-\mathbf{\Delta}$ [11].

3.2 Computing Amplitude Spectra on Very Large Graphs

For graphs with more than a few thousand vertices it becomes impossible to compute all eigenvectors directly (e.g. using the QR algorithm) since the Laplacian matrix $-\mathbf{\Delta}$ becomes too large even to fit into the computer's memory. However, in many cases these matrices are extremely sparse. Moreover, we may know the Laplacian eigenvalues from its equitable partitions. In this case we use the accompanying knowledge of the spectral radius of the Laplacian to transform the problem to a more tractable eigenspace computation.

Lemma 1 *Let* $\mathbf{B}$ *be a matrix with spectral radius* $\varrho(\mathbf{B})$ *let* ξ *be an eigenvector of* $\mathbf{B}$ *belonging to the eigenvalue* λ. *Then the matrix* $\mathbf{Q}_\lambda$, *defined by*

$$\mathbf{Q}_\lambda = \mathbf{I} - \frac{1}{2\,(1+\varrho(\mathbf{B})^2)}(\mathbf{B}-\lambda\mathbf{I})^2 \tag{17}$$

has spectral radius $\varrho(\mathbf{Q}_\lambda) = 1$ *and* $\|\mathbf{Q}_\lambda\xi\| = \|\xi\|$ *if and only if* $\mathbf{B}\xi = \lambda\xi$.

Proof. Consider an eigenvector ζ of $\mathbf{B}$ with eigenvalue μ. We have

$$1 \geq \frac{\langle\zeta, \mathbf{Q}_\lambda\zeta\rangle}{\langle\zeta,\zeta\rangle} = 1 - \frac{(\mu-\lambda)^2}{2(\varrho(\mathbf{B})^2+1)} \geq 1 - \frac{2\varrho(\mathbf{B})^2}{1+\varrho(\mathbf{B})^2} > -1$$

We see immediately that $\langle\zeta, \mathbf{Q}\zeta\rangle \leq \langle\zeta,\zeta\rangle$ with equality if and only if $\mu = \lambda$. The lemma now follows directly from the properties of the Rayleigh quotient. $\square$

Lemma 1 allows us to obtain the amplitude of a landscape f within the eigenspace of $\mathbf{B}$ belonging to λ from $\mathbf{Q}_\lambda$ by an iterative procedure.

Lemma 2 *Let* f *be a landscape satisfying (11) and let* λ *be an eigenvalue of* $\mathbf{B}$. *Then the amplitude* $B(\lambda)$ *of* f *within the eigenspace of* $\mathbf{B}$ *belonging to* λ

satisfies

$$B(\lambda) = \lim_{k\to\infty} \langle f, \mathbf{Q}_\lambda^k f\rangle \tag{18}$$

Proof. Let $\{\vartheta_k\}$ be an orthonormal basis of eigenvectors of $\mathbf{Q}$ with associated eigenvalues μ_k. We expand $f = \sum_k a_k \vartheta_k$ and obtain

$$\langle f, \mathbf{Q}_\lambda^k f\rangle = \sum_{i,j} a_i \mu_j^k a_j \langle \vartheta_i, \vartheta_j\rangle = \sum_i a_i^2 \mu_i^k$$

Lemma 1 now implies $\sum_{i:\mu_i=1} a_i^2 = B(\lambda)$ because the eigenvectors of $\mathbf{Q}_\lambda$ with eigenvalues $\mu_k = 1$ span the eigenspace of $\mathbf{B}$ with eigenvalue λ. $\square$

The utility of (18) derives from the fact that by the definition of $\mathbf{Q}_\lambda$ (17) we may compute $\mathbf{Q}_\lambda f$ by instead computing $\mathbf{B}f$ and $\mathbf{B}^2 f$ without storing more than the "adjacency list" of the graph. If we use β_x to denote the set of vertices adjacent to x we obtain

$$[\mathbf{B}f]_x = \sum_{y\in\beta_x} f(y) \qquad \text{and} \qquad [\mathbf{B}^2 f]_x = \sum_{y\in\beta_x}\sum_{z\in\beta_y} f(z) \tag{19}$$

On the other hand (18) is inconvenient for practical computations since inaccuracies in $\mathbf{Q}_\lambda$ are amplified exponentially. Thus we first iterate $f^{(k+1)} = (1/\|\mathbf{Q}_\lambda f^{(k)}\|)\mathbf{Q}_\lambda f^{(k)}$ to convergence and compute $B(\lambda) = \langle f, f^{(\infty)}\rangle$.

If $B(\lambda) = 0$, or is very small, then it is possible that $f^{(k)}$ can converge to an eigenvector of $\mathbf{Q}_\lambda$ whose eigenvalue associated eigenvalue is smaller than 1, but whose projection onto the eigenspace is still non-vanishing. Hence it is necessary to check that $\|f^{(\infty)} - \mathbf{Q}_\lambda f^{(\infty)}\|$ is smaller than a prescribed tolerance. Otherwise we have $B(\lambda) = 0$ (within the numerical tolerance).

The major drawback of this approach is that the speed of convergence implicit in (18) can be very poor. The error term is given by the contributions of all other eigenspaces and hence is bounded above by $\varepsilon(k) = \mu_2^k$ after k iterations, where μ_2 denotes the second-largest eigenvalue of $\mathbf{Q}_\lambda$. From (17) we have

$$\mu_2 \approx 1 - \frac{1}{2}\left(\frac{\Delta\lambda}{\varrho(\mathbf{B})}\right)^2, \tag{20}$$

where $\Delta\lambda$ is the minimum difference between λ and the next closest eigenvalue of $\mathbf{B}$. The number of iteration required to reach an accuracy of K digits therefore scales as $\mathcal{O}\left[K \cdot (\varrho(\mathbf{B})/\Delta\lambda)^2\right]$. A more efficient numerical procedure would therefore be desirable if the Laplacian spectrum contains eigenvalues which are not well separated. We remark that a straightforward application of a Lanczos iteration does not work because of the large multiplicities of most eigenvalues [23].

Local optima offer an alternative approach to quantify ruggedness. Intuitively, a more rugged landscape implies more local optima. It suffices to consider only local minima.

Definition 2 *A vertex x is a* local minimum *of a landscape f on $\Gamma = (X, \mathcal{N})$ if $f(x) \leq f(y)$ for all neighbors y of x.*

The use of $\leq$ instead of $<$ is customary in discrete systems.

Local optima are separated by saddle points and *fitness barriers*. Let $\hat{x}$ and $\hat{y}$ be two local minima and let $\mathbf{p}$ be a path in Γ from $\hat{x}$ to $\hat{y}$. Then the fitness barrier separating $\hat{x}$ from $\hat{y}$ is

$$f[\hat{x}, \hat{y}] = \min \left\{ \max \left[f(z) \middle| z \in \mathbf{p} \right] \middle| \mathbf{p} : \text{path from } \hat{x} \text{ to } \hat{y} \right\} \tag{21}$$

A point $\hat{z} \in X$ satisfying the minimax condition (21) is a *saddle point* of the landscape. The saddle point fitnesses $f[\hat{x}, \hat{y}]$ form an ultrametric distance measure on the set of local minima (see e.g. [38]). This hierarchical structure can be represented by the *barrier tree* of the landscape. Its leaves are the local minima and its internal nodes correspond to saddle points.

Barrier trees have been considered recently for various models of disordered systems ranging from biopolymers to spin glasses and combinatorial optimization [5, 18, 20, 22, 32]. In this contribution we use a modified version of the program `barriers`, which was originally developed for the analysis of RNA folding landscapes [20]. A detailed description of the algorithm can be found in [18], where `barriers` is used to analyze the energy landscapes of spin glass models.

Figure 5 shows the amplitude spectrum and barrier tree for an uncorrelated random landscape on Γ_7. The *barrier* enclosing a local minimum is the height of the lowest saddle point that gives access to a more favorable minimum. In symbols:

$$B(\hat{x}) = \min \left\{ f[\hat{x}, \hat{y}] - f(\hat{x}) \middle| \hat{y} : \text{local minimum such that } f(\hat{y}) < f(\hat{x}) \right\} \tag{22}$$

The barrier heights can be directly read off the barrier trees. We use the middle panel of Figure 6 as an example: $B(3) = 0.63$, the fitness difference between the local minimum of 3 and the saddle point separating 3 from 2. The value $B(2) = 1.62$ is the fitness of the saddle point separating 1 from the subtree $\{2, 3, 4, 5, 6\}$.

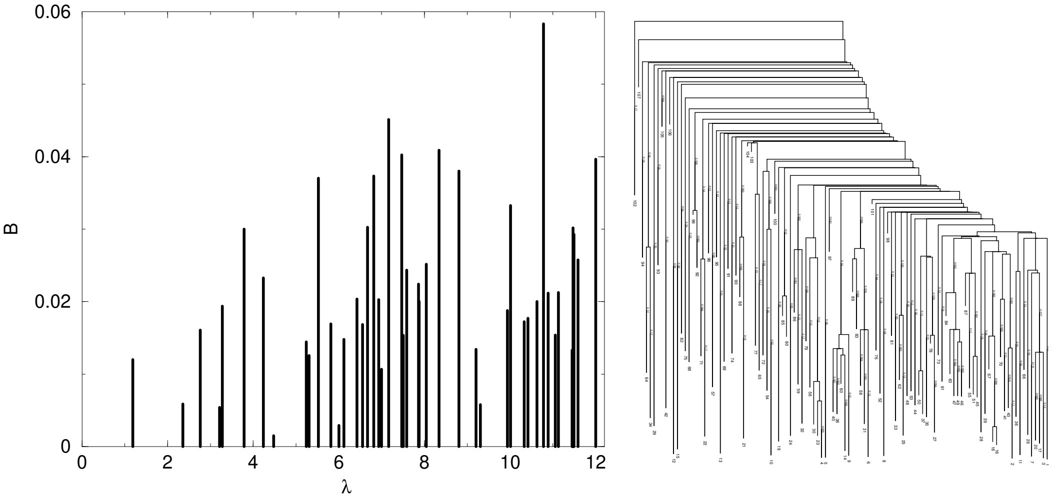

Fig. 5. Amplitude spectrum and barrier tree of a REM on Γ_7.

If $B(\hat{x}) = 0$ then the local minimum $\hat{x}$ is marginally stable. It is easy to check that eq.(22) is equivalent to the definition of the *depth* of a local minimum in [30]. It agrees for meta-stable states with the more general definition of the depth of a "cycle" in the literature on inhomogeneous Markov chains [1, 8, 9]. The barrier height $B(\hat{x})$ essentially measures how hard it is to escape form a local minimum.

4 Landscapes on Robinson Graphs

4.1 Simple Examples

It will be convenient to represent a phylogenetic tree by the collection of its *splits*: Let $e \in E(T)$ is an edge in the phylogenetic tree T, and denote by L^e and $\bar{L}^e$ are the sets of leaves in the two connected components of $T \setminus \{e\}$. The pair $s_e^T = \{L^e, \bar{L}^e\}$ is the split associated with the edge e. A set of splits $\mathcal{S}(T) = \{s_e^T | e \in T\}$ is *compatible* if for any two splits $\{A, A'\}, \{B, B'\} \in \mathcal{S}(T)$ at least one of the four intersections $A \cap B$, $A \cap B'$, $A' \cap B$, $A' \cap B'$ is empty. The split set of tree T is always compatible, and conversely, a compatible system of splits defines a unique tree [7].

Proposition 4 [47, 40] *Two phylogenetic trees T and T' are* nni *neighbors if and only if the symmetric difference $\mathcal{S}(T) \triangle \mathcal{S}(T')$ contains exactly* 2 *splits.*

In the following let $v \in \hat{V}(T)$ be an interior vertex of T. The three edges incident with v will be denoted by e_v, e'_v, e''_v. The close relationship between splits and the nni neighborhood suggests that simple landscapes on Robinson

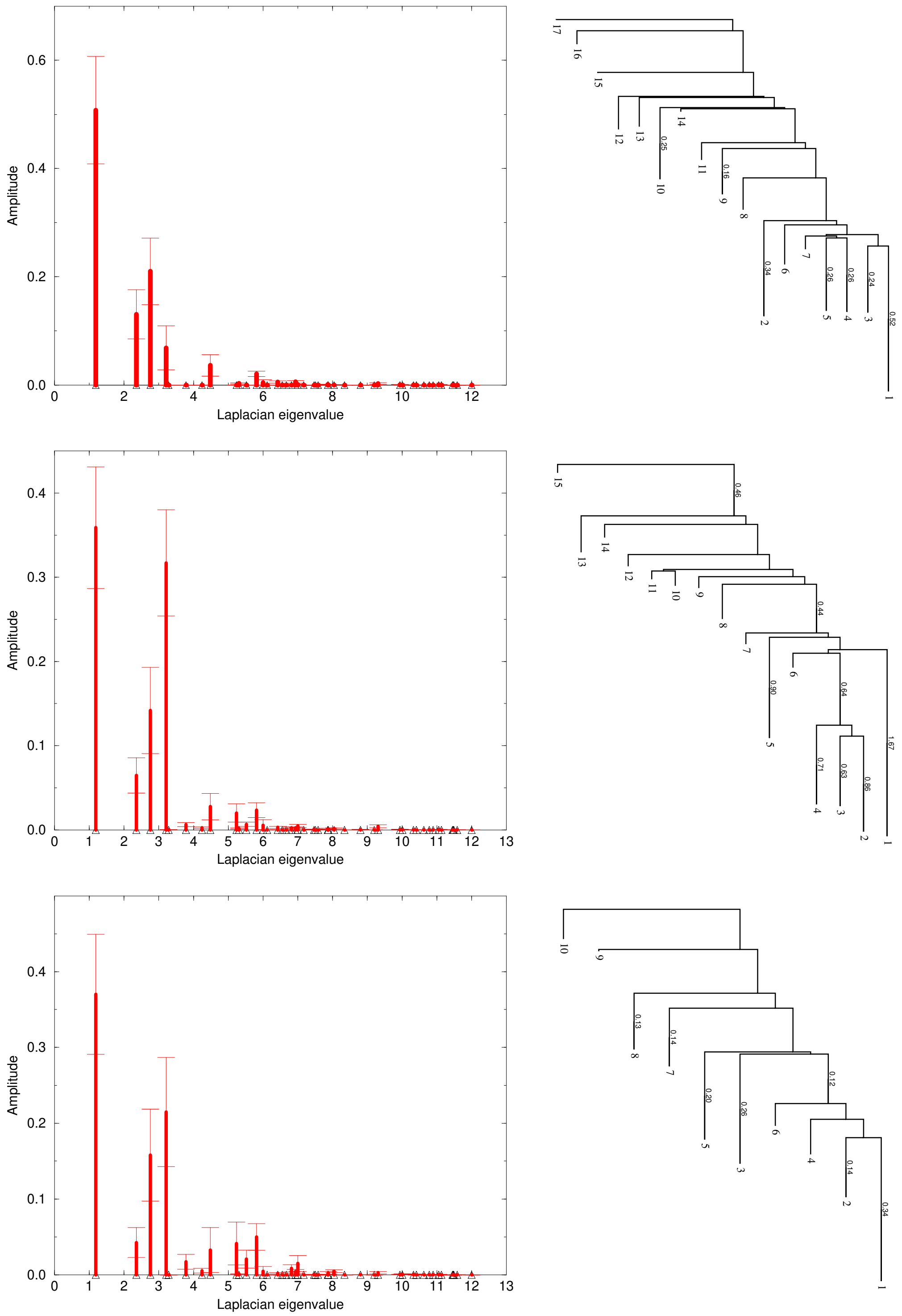

Fig. 6. Expected Amplitude spectra (l.h.s.) and barrier trees (r.h.s) of the simple split-based cost functions defined (from top to bottom) in equations (23-25) for $n = 7$. The error bars show the standard deviations of the distribution of amplitudes when the split weights $\mu_i(s)$ are drawn independently from a uniform distribution.

graphs can be constructed as sums of weights $\mu_i(s)$ defined for the individual splits. We consider the following three simple cost functions:

$$
\begin{aligned}
f_1(T) &= \sum_{s\in\mathcal{S}(T)} \mu_1(s) && (23)\\
f_2(T) &= \sum_{s\in\mathcal{S}(T)} \Big[\mu_2(s^T_{e_v})\mu_2(s^T_{e'_v}) + \mu_2(s^T_{e_v})\mu_2(s^T_{e''_v}) + \mu_2(s^T_{e'_v})\mu_2(s^T_{e''_v})\Big] && (24)\\
f_3(T) &= \sum_{v\in\hat{V}(T)} \mu_3(s^T_{e_v})\mu_3(s^T_{e'_v})\mu_3(s^T_{e''_v}) && (25)
\end{aligned}
$$

The examples shown in Figure 6, the weights $\mu_i(s)$ drawn independently from a uniform distribution. Each of these three landscapes has an amplitude spectrum that is concentrated around the lowest non-zero eigenvalue Λ_1 of the Laplacian. Their barrier trees are also similar. Not surprisingly, the number of local minima is much smaller than in the uncorrelated random landscape of Figure 5. Also note that most local minima have low barriers.

4.2 The Maximum Parsimony Problem

Suppose we are given a matrix $C = (c_i(v))$ describing m characters of a set of n species. In molecular systematics, the m characters are the m positions of an DNA or protein sequence of length m, or more precisely, C is a multiple alignment of n sequences, one from each species (see e.g. [46]). What is the optimal phylogeny for these species, i.e., the one minimizing the number of mutation events?

Given a tree T and the vector of m characters for each vertex v of T the *parsimony score* of $\mathrm{ps}(T)$ is defined as

$$
\mathrm{ps}(T) = \sum_{(u,v)\in E(T)} \Big|\{j : c_j(u) \neq c_j(v)\}\Big| \qquad (26)
$$

In the context of molecular biology $\mathrm{ps}(T)$ is the total number of mutation events along the phylogenetic tree. In practice, however, the characters $c_i(v)$ are only known for the leaves of T, not for the internal vertices. The *small parsimony problem* consists of finding an assignment of characters to the interior nodes of T such that $\mathrm{ps}(T)$ is minimized, subject to a given set of characters at the leaves. This problem can be solved in $\mathcal{O}(n)$ steps using Fitch's algorithm [19] for a single character $c(v)$ at each leave. The small parsimony problem is therefore solved by applying Fitch's algorithm separately for each character. The resulting total score can be interpreted as the minimum number of mutation events that can generate the observed characters at the leaves *given*

Table 2
The largest amplitudes of the parsimony landscapes of Figure 7.

p	Λ_p	random	aligned random	Picorna
1	0.987478	**0.7609**	**0.8325**	**0.7359**
4	2.362683	**0.0949**	**0.0956**	**0.1889**
5	2.670694	0.0075	**0.0134**	**0.0220**
8	3.244050	0.0045	0.0046	**0.0160**
12	4.083940	0.0015	0.0055	**0.0104**
20	5.000000	**0.0318**	0.0002	0.0006
35	6.000000	**0.0147**	0.0002	0.0001

the order of the speciation events (i.e., given the tree T). It will be denoted by $p_c(T)$ to emphasize that it depends on the input data c. The maximum parsimony problem consists of finding the tree that minimizes $p_c(T)$ over the set of all trees $T \in \mathfrak{B}_n$. Maximum parsimony is a widely used technique for inferring molecular phylogenies [17].

Maddison [35] noted that heuristic search procedures for maximum parsimony trees as implemented in programs such as `PAUP` or `Henning86` can get stuck in multiple local optima ("tree islands"). This study focussed on the `tbr` move set, since this implies the same phenomenon for the more restricted move sets `spr` and `nni`. Maddison reports that multiple local optima are likely for `tbr` when the so-called retention index [15] is small.

As a final example we compare here the parsimony score landscapes of three different datasets consisting of $n = 8$ sequences each: (a) random sequences of length 100, (b) aligned random sequences (using `ClustalW`) of length 100, and (c) an alignment of eight complete Picorna virus genomes. For each case we compute both the amplitude spectrum and the the barrier trees (see Figure 7). A comparison of both the amplitude spectra and the barrier trees of the parsimony landscapes in Figure 7 with the simple models of the previous section shows that the parsimony landscape is exceptionally smooth. About 75% of the variance is contained in the eigenspace of the smallest non-zero eigenvalue Λ_1. In table 2 we list those eigenspaces which carry at least 1% of the variance in one of the three parsimony score landscapes. The landscapes arising from properly aligned sequences are significantly smoother than the random sequence version, as can be seen both from the barrier trees and the amplitude spectrum: the random sequences lead to significant contributions from large eigenvalues, Λ_{20} and Λ_{35} which are virtually absent from the aligned versions.

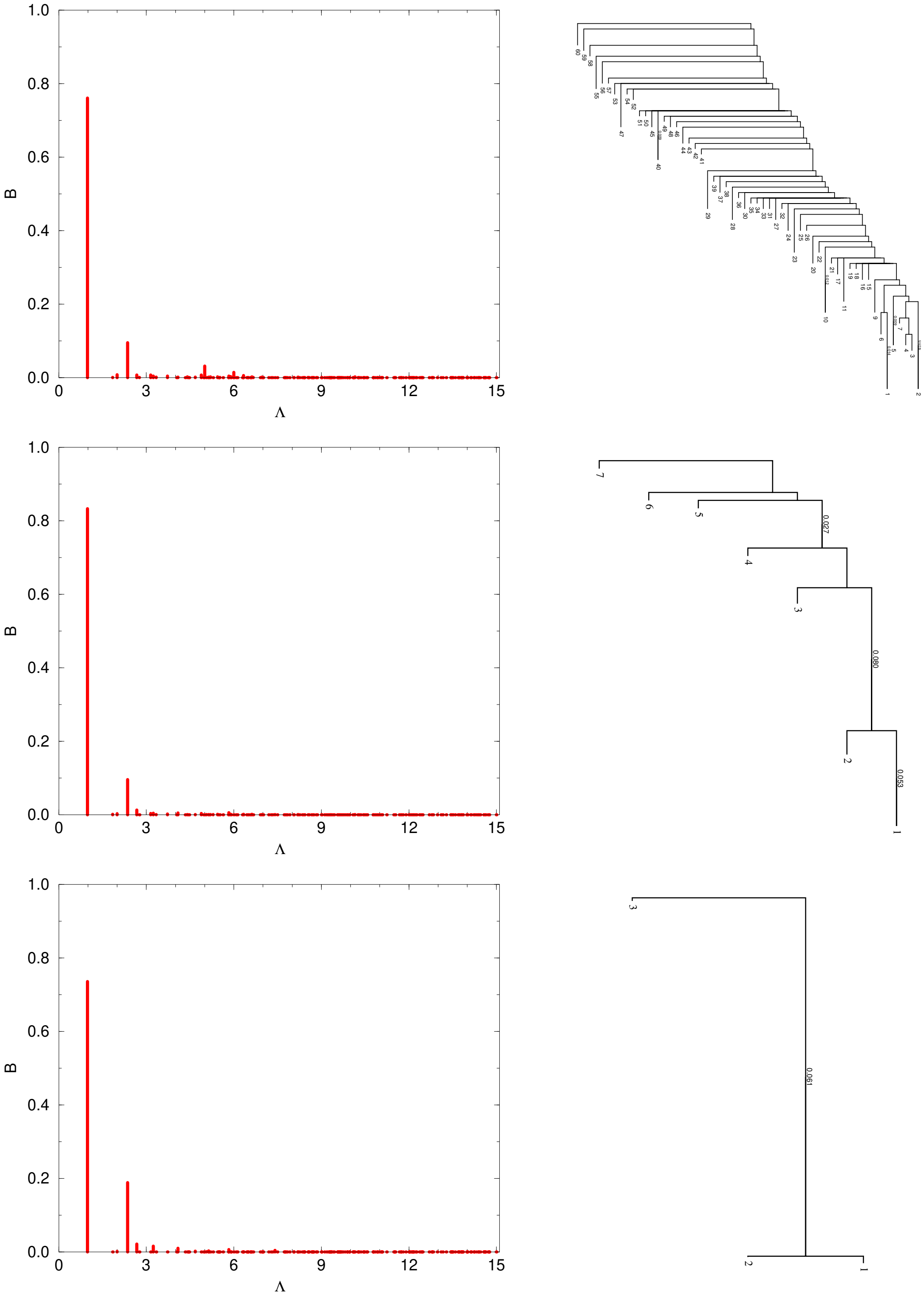

Fig. 7. Parsimony Landscapes. Top row: Parsimony score from random sequences. Middle row: Parsimony score from `Clustal W` aligned random sequences, Bottom row: Parsimony score from an alignment of eight Picorna Virus genomes.

Acknowledgments

This work was supported in part by NSF, PFF Award DMS-9553134, and the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung*, Proj. No. P14094-MAT. Thanks to Christoph Flamm for the source code of `barriers`.

References

[1] R. Azencott. *Simulated Annealing.* John Wiley & Sons, New York, 1992.

[2] O. Bastert. New ideas for canonically computing graph algebras. Technical Report TUM-M9803, Technische Universität München, München, Germany, 1998.

[3] O. Bastert. Computing equitable partitions of graphs. *MATCH*, 40:265–272, 1999.

[4] O. Bastert, D. Rockmore, P. F. Stadler, and G. Tinhofer. Some properties of Robinson graphs. Technical Report TUM-M0081, Technical University Munich, 2000.

[5] O. M. Becker and M. Karplus. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Chem. Phys.*, 106:1495–1517, 1997.

[6] A. Brouwer, A. Cohen, and A. Neumaier. *Distance-regular Graphs.* Springer Verlag, Berlin, New York, 1989.

[7] P. Buneman. The recovery of trees from measures of dissimilarity. In F. R. Hodson, D. G. Kendall, and P. Tautu, editors, *Mathematics and the Archeological and Historical Sciences*, pages 387–395. Edinburgh University Press, Edinburgh, UK, 1971.

[8] O. Catoni. Rough large deviation estimates for simulated annealing: Application to exponential schedules. *Ann. Probab.*, 20:1109–1146, 1992.

[9] O. Catoni. Simulated annealing algorithms and Markov chains with rate transitions. In J. Azema, M. Emery, M. Ledoux, and M. Yor, editors, *Seminaire de Probabilites XXXIII*, volume 709 of *Lecture Notes in Mathematics*, pages 69–119. Springer, Berlin/Heidelberg, 1999.

[10] F. R. K. Chung. *Spectral Graph Theory*, volume 92 of *CBMS*. American Mathematical Society, 1997.

[11] E. B. Davies, G. M. L. Gladwell, J. Leydold, and P. F. Stadler. Discrete nodal domain theorems. *Lin. Alg. Appl.*, 2001. See also `xxx.lanl.gov math.SP/0009120`.

[12] W. H. E. Day, D. S. Johnson, and D. Sankoff. The computational complexity of inferring rooted phylogenies by parsimony. *Math. Biosci.*, 81:33–42, 1986.

[13] I. A. Faradžev, M. H. Klin, and M. E. Muzychuk. Cellular rings and groups of automorphisms of graphs. In I. A. Faradžev, A. A. Ivanov,

M. H. Klin, and A. J. Woldar, editors, *Investigations in Algebraic Theory of Combinatorial Objects*, volume 84 of *Mathematics and Its Applications Soviet Series*. Kluwer, Dordrecht, 1994.

[14] J. S. Farris. The logical basis of phylogenetic analysis. In N. I. Platnick and V. A. Funk, editors, *Advances in Cladistics*, pages 1–36. Columbia University Press, New York, 1983.

[15] J. S. Farris. The retention index and the rescaled consistency index. *Cladistics*, 5:417–419, 1989.

[16] J. Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, 17:368–376, 1981.

[17] J. Felsenstein. Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.*, 22:368–376, 1988.

[18] F. F. Ferreira, J. F. Fontanari, and P. F. Stadler. Landscape statistics of the low autocorrelated binary string problem. *J. Phys. A: Math. Gen.*, 33:8635–8647, 2000.

[19] W. M. Fitch. Toward defining the course of evolution: minimum change for a specified tree topology. *Systematic Zoology*, 20:406–416, 1971.

[20] C. Flamm, W. Fontana, I. Hofacker, and P. Schuster. RNA folding kinetics at elementary step resolution. *RNA*, 6:325–338, 2000.

[21] L. R. Foulds and R. L. Graham. The Steiner problem in phylogeny is NP-complete. *Adv. Appl. Math.*, 3:43–49, 1982.

[22] P. Garstecki, T. X. Hoang, and M. Cieplak. Energy landscapes, supergraphs, and "folding funnels" in spin systems. *Phys. Rev. E*, 60:3219–3226, 1999.

[23] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins Univ. Press, Balimore, MD, 1983.

[24] L. K. Grover. Local search and the local structure of NP-complete problems. *Oper. Res. Lett.*, 12:235–243, 1992.

[25] R. Happel and P. F. Stadler. Canonical approximation of fitness landscapes. *Complexity*, 2:53–58, 1996.

[26] D. Higman. Coherent configurations. I: Ordinary representation theory. *Geometriae Dedicata*, 4:1–32, 1975.

[27] D. Higman. Coherent configurations. Part II: Weights. *Geometriae Dedicata*, 5:413–424, 1976.

[28] D. Higman. Coherent algebras. *Lin. Alg. Appl.*, 93:209–239, 1987.

[29] W. Hordijk and P. F. Stadler. Amplitude spectra of fitness landscapes. *J. Complex Systems*, 1:39–66, 1998.

[30] W. Kern. On the depth of combinatorial optimization problems. *Discr. Appl. Math.*, 43:115–129, 1993.

[31] M. Klin, C. Rücker, G. Rücker, and G. Tinhofer. Algebraic combinatorics in mathematical chemistry. Methods and algorithms. I. Permutation groups and coherent (cellular) algrebras. *MATCH*, 40:7–138, 1999.

[32] T. Klotz and S. Kobe. "valley structures" in the phase space of a finite 3D Ising spin glass with $\pm i$ interactions. *J. Phys. A: Math. Gen*, 27:L95–L100, 1994.

[33] J. Lafferty and D. Rockmore. Numerical investigation of the spectrum for certain families of Cayley graphs. In J. Friedman, editor, *Expanding Graphs*, volume 10 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 63–73. American Mathematical Society, 1993.

[34] M. Li, J. Tromp, and L. Zhang. On the nearest neighbour interchange distance between evolutionary trees. *J. Theor. Biol.*, 182:463–467, 1996.

[35] D. R. Maddison. The discovery and importance of multiple islands of most parsimonious trees. *Systematic Zoology*, 40:315–328, 1991.

[36] R. Merris. A survey of graph Laplacians. *Lin. Alg. Appl.*, 39:19–31, 1995.

[37] B. Mohar. Some applications of Laplace eigenvalues of graphs. In G. Hahn and G. Sabidussi, editors, *Graph Symmetry: Algebraic Methods and Applications*, volume 497 of *NATO ASI Series C*, pages 227–275. Kluwer, Dordrecht, 1997.

[38] R. Rammal, G. Toulouse, and M. A. Virasoro. Ultrametricity for physicists. *Rev. Mod. Phys.*, 58:765–788, 1986.

[39] D. F. Robinson. Comparison of labeled trees with valency three. *J. Combin. Theory B*, 11:105–119, 1971.

[40] D. F. Robinson and L. R. Foulds. Comparison of phylogenetic trees. *Math. Biosci.*, 53:131–147, 1981.

[41] E. Schröder. Vier Combinatorische Probleme. *Zeitschr. f. Math. Phys.*, 15:361–376, 1870.

[42] P. F. Stadler. Landscapes and their correlation functions. *J. Math. Chem.*, 20:1–45, 1996.

[43] P. F. Stadler. Spectral landscape theory. In J. P. Crutchfield and P. Schuster, editors, *Evolutionary Dynamics—Exploring the Interplay of Selection, Neutrality, Accident, and Function.* Oxford University Press, 2000. in press.

[44] P. F. Stadler and G. Tinhofer. Equitable partitions, coherent algebras and random walks: Applications to the correlation structure of landscapes. *MATCH*, 40:215–261, 1999.

[45] D. L. Swofford and G. J. Olsen. Phylogeny reconstruction. In D. M. Hillis and C. Moritz, editors, *Molecular Systematics*, pages 411–501. Sinauer Associates, Sunderland MA, 1990.

[46] M. S. Waterman. *Introduction to computational biology: maps, sequences and genomes.* Chapman & Hall, London, New York NY, 1995.

[47] M. S. Waterman and T. F. Smith. On the similarity of dendrograms. *J. Theor. Biol.*, 73:789–800, 1978.

[48] E. D. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biol. Cybern.*, 63:325–336, 1990.

[49] B. Y. Weisfeiler, editor. *On Construction and Identification of Graphs*, volume 558 of *Lecture Notes in Mathematics*. Springer, Berlin, 1976.

[50] B. Y. Weisfeiler and A. A. Leman. Reduction of a graph to a canonical form and an algebra arising during this reduction. *Naucho – Techn. Inf.; Ser. 2*, 9:12–16, 1968. Russian.