Computers and Mathematics with Applications 62 (2011) 3821–3828

Contents lists available at SciVerse ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

Computing quadratic entropy in evolutionary trees

Drago Bokal [a,*,1], Matt DeVos [b], Sandi Klavžar [c,a,1], Aki Mimoto [d,2], Arne Ø. Mooers [d,2]

[a] Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, SI-2000 Maribor, Slovenia

[b] Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby BC V5A 1S6, Canada

[c] Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

[d] Department of Biological Sciences, Simon Fraser University, 8888 University Drive, Burnaby BC V5A 1S6, Canada

ARTICLE INFO

Article history:
Received 25 March 2011
Received in revised form 14 September 2011
Accepted 15 September 2011

Keywords:
Evolutionary tree
Phylogenetic tree
Quadratic entropy
Originality
Distinctness
Wiener index

ABSTRACT

We note here that quadratic entropy, a measure of biological diversity introduced by C.R. Rao, is a variant of the weighted Wiener index, a graph invariant intensively studied in mathematical chemistry. This fact allows us to deduce some efficient algorithms for computing the quadratic entropy in the case of given tip weights, which may be useful for community biodiversity measures. Furthermore, on ultrametric phylogenetic trees, the maximum of quadratic entropy is a measure of pairwise evolutionary distinctness in conservation biology, introduced by S. Pavoine. We present an algorithm that maximizes this quantity in linear time, offering a significant improvement over the currently used quadratic programming approaches.

1. Introduction

Phylogenetic trees are simply graphs depicting the inferred relationships among predefined sets of leaves (which often correspond to species). This means that they are amenable to analyses with graph theory [1]. If they are given a direction by identifying a root, we can speak about *evolutionary trees*. Their structure models evolution, which has a direction from past to present, and which is generally (but not exclusively, see [2]) diversifying, and such that the simultaneous production of more than two descendant lineages from an ancestral lineage is rare. Biologists often consider internal vertices to represent extinct ancestral lineages and the edge lengths to represent the amount of evolution that occurred between the species corresponding to the endvertices. Evolutionary trees are most often inferred by fitting an evolutionary process model to discrete data measured on the leaves in a maximum likelihood or Bayesian framework [3]. Because evolutionary trees are representations of the evolutionary history of a set of contemporaneous leaves, they are often forced to be *ultrametric*, i.e. all leaves are equidistant from the root. Such an evolutionary tree has a height h, the sum of edge lengths on the path from the root to (any) leaf; edge lengths are then inferred to represent the relative elapsed time between internal vertices.

Past mathematical research has considered tree inference, tree shape distribution and accompanying generating models, as well as parameter estimation from inferred trees. So, Semple and Steel [1] summarize how graph theory can contribute to the NP-complete problem of evolutionary tree inference, e.g. what the mathematical properties allow for recovery of the underlying tree under different assumptions concerning character evolution, and how subtrees can be combined to

* Corresponding author.
E-mail address: drago.bokal@uni-mb.si (D. Bokal).

1 Also with: Institute of Mathematics, Physics and Mechanics, Jadranska 19, 1000 Ljubljana, Slovenia.

2 Also with: Interdisciplinary Research in Computational and Mathematical Sciences Center, Simon Fraser University, Canada.

0898-1221/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2011.09.030

best preserve their information. Explorations of evolutionary tree structure distributions have a long pedigree, with most attention focused on the Yule [4] and uniform [5] distributions of topologies [6,7]. There has also been related discussion on appropriate prior distributions (of tree topology and edge lengths) for evolutionary tree inference [8–10], and efficient algorithms listing all possible evolutionary trees for a given set of species have been developed [11]. At the other end of the evolutionary tree inference cycle, mathematically-inclined biologists have produced tools for estimating evolutionary parameters (speciation and extinction rates) from inferred trees [12–14].

Graph theory can also bear on practical biological conservation. If we assume that one aspect of the leaves (species) that humans would like to conserve is the unique information they embody, and if we let the edge-weighted evolutionary trees represent the pattern of shared and unique information, we can start to devise approaches that maximize this quantity (called 'evolutionary history' or 'phylogenetic diversity') under constraints of final subset size, budgets for conservation, costs of conservation, and the probabilities of species survival [15,16]. This has been termed the "Noah's Ark Problem" [17]. A related metric is the contribution of a leaf to future subsets on an evolutionary tree—these have been termed a leaf's 'originality' or 'distinctness' [18–24].

In this contribution, we explore one such measure, *quadratic entropy*, introduced by Rao [25] and recently applied to evolutionary trees by Pavoine [22,26]. These authors propose computing the probability distribution μ maximizing the quadratic entropy for general finite metric spaces. In the evolutionary context, Pavoine [22] specifically suggests using μ as an importance score for species on a tree as a weight representing its expected pairwise contribution of evolutionary originality. The method can be interpreted as finding an optimum of a quadratic mathematical program, yielding an algorithm of complexity $O(n^4)$. It has been implemented as a function [27] in the ADE package for analysis of environmental data [28] within the statistical environment *R* [29]. However, one can use the specific structure of evolutionary trees to develop a linear time algorithm for maximizing the quadratic entropy in two depth-first traversals of the tree. Presenting this algorithm (implemented in R [30]) is the main goal of the present contribution. We also make a few additional observations on computing the quadratic entropy and its connections with the graph invariant Wiener index [31–34], widely known in mathematical chemistry. This last connection also allows for an alternative and rapid (linear time) algorithm of computing the quadratic entropy on evolutionary trees with known leaf weights.

2. Evolutionary trees and quadratic entropy

An *evolutionary tree* $\mathcal{T} = (T, r, w)$ consists of a tree T rooted at a vertex $r \in V(T)$ whose edges have their length determined by a function $w : E(T) \to \mathbb{R}^+ \cup \{0\}$. Between any two pairs of vertices $u, v \in V(T)$, there is a unique shortest path in T, and by the *distance* $d(u, v)$ between u and v we denote the *length* of this path, i.e. the sum of the w-values of its edges. In a rooted tree, every vertex $v \in V(T)$ has a unique incident edge e_v that lies on the shortest path connecting v with r. The component of $T - e_v$ containing v is the *subtree* T_v *rooted at* v. Then, $\mathcal{T}_v = (T_v, v, w/_{E(T_v)})$ is the corresponding evolutionary subtree. The endvertex of e_v distinct from v, is the *parent* of v, and all neighbors for which v is a parent are *children* of v.

Each vertex in an evolutionary tree represents a species in the history of Earth. A leaf vertex represents either a living species or an extinct species, and an internal vertex represents the common ancestral species of those corresponding to the vertices in $V(T_v)$. The length of an edge uv represents the time elapsed between the species v, whose immediate ancestor is u, branched into two or (rarely) more new species. Therefore the living species are represented by the leaves at the largest distance h from r, called the *height* of $\mathcal{T}$. We assume that $\mathcal{T}$ contains no extinct leaf species, i.e. all the leaves of $\mathcal{T}$ are at distance h from r, making $\mathcal{T}$ *strictly ultrametric*. The height corresponds to the age of the species represented by r. Note that, in the case that the root r has degree one, we do not consider it as a leaf of $\mathcal{T}$.

Let μ be a probability distribution on the leaves of $\mathcal{T}$ and let D denote the random variable, representing the distance among two μ-randomly selected leaves of $\mathcal{T}$ (with repetition). *Quadratic entropy* $\mathbb{E}(D)$ is the expected value of this random variable [22,26,25]. We can define it for any metric space X, in which case we are evaluating the expected distance between two randomly selected elements of X. Thus, if μ is the vector of relative frequencies of elements from X (in our case, species) and A is the corresponding distance matrix (in our case, the matrix of distances in the evolutionary tree), then

$$\mathbb{E}(D) = \mu^T A \mu.$$

Matrix multiplication from this formula yields a quadratic algorithm for computing $\mathbb{E}(D)$ for given μ and A. Further, finding the probability distribution μ that maximizes $\mathbb{E}(D)$ for given A corresponds to finding a maximum of a quadratic program in variable μ and can be done using standard methods of convex programming. In the special case of evolutionary trees, we develop significantly more efficient algorithms for both tasks, which run in linear time.

3. Computing quadratic entropy

Suppose $\mathcal{T}_1 = (T_1, r_1, w_1)$ and $\mathcal{T}_2 = (T_2, r_2, w_2)$ are two evolutionary trees. The *join* of these two trees is the tree $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2, \mathcal{T} = (T, r, w)$, where T is obtained from T_1 and T_2 by identifying their roots r_1 and r_2 into a new root r, and the length function w of $\mathcal{T}$ is induced by the functions w_1 and w_2. For a binary tree, At each node the operation is performed only once, but several iterative applications (at each internal vertex one less than the number of children) can be combined to construct an arbitrary tree, cf. Fig. 1.

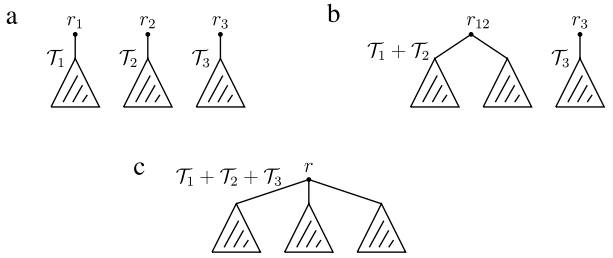

Fig. 1. The iterative join of evolutionary trees $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$.

For a subtree $\mathcal{T}'$ of $\mathcal{T}$, let $\mathbb{E}(D \mid \mathcal{T}')$ denote the expected value of D conditional to both leaves being selected from $\mathcal{T}'$, and let $\mu(\mathcal{T}')$ denote the sum of $\mu(l)$ for all leaves l in $\mathcal{T}'$, i.e. the probability that a random leaf of $\mathcal{T}$ is a leaf of $\mathcal{T}'$. Our key observation, made precise in Proposition 1, is, that for an ultrametric tree $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2$, the expected value $E(D)$ for $\mathcal{T}$ can be iteratively computed from $E(D \mid \mathcal{T}_1)$ and $E(D \mid \mathcal{T}_2)$, where we use the fact that any leaf from $\mathcal{T}_1$ is at the same distance to any leaf in $\mathcal{T}_2$. A similar argument applies to maximizing quadratic entropy, thus both computation algorithms follow the manner in which a tree is constructed by iteratively joining its subtrees, cf. Fig. 1. The following proposition holds:

Proposition 1. *Let $\mathcal{T}$ be a tree with probability distribution μ on its leaves, and let $\mathcal{T}$ be the join of two trees $\mathcal{T}_1$ and $\mathcal{T}_2$ of the same height h. Then,*

$$\mathbb{E}(D) = \mathbb{E}(D \mid \mathcal{T}_1)\mu(\mathcal{T}_1)^2 + \mathbb{E}(D \mid \mathcal{T}_2)\mu(\mathcal{T}_2)^2 + 4h\mu(\mathcal{T}_1)\mu(\mathcal{T}_2).$$

Proof.

$$\begin{aligned}
\mathbb{E}(D) &= \sum_{l,l' \in \mathcal{T}} \mu(l)\mu(l')d(l,l') \\
&= \sum_{l,l' \in \mathcal{T}_1} \mu(l)\mu(l')d(l,l') + \sum_{l,l' \in \mathcal{T}_2} \mu(l)\mu(l')d(l,l') + \sum_{l \in \mathcal{T}_1, l' \in \mathcal{T}_2} \mu(l)\mu(l')d(l,l') + \sum_{l \in \mathcal{T}_2, l' \in \mathcal{T}_1} \mu(l)\mu(l')d(l,l') \\
&= \mathbb{E}(D \mid \mathcal{T}_1)\mu(\mathcal{T}_1)^2 + \mathbb{E}(D \mid \mathcal{T}_2)\mu(\mathcal{T}_2)^2 + 4h \sum_{l \in \mathcal{T}_1} \mu(l) \sum_{l' \in \mathcal{T}_2} \mu(l') \\
&= \mathbb{E}(D \mid \mathcal{T}_1)\mu(\mathcal{T}_1)^2 + \mathbb{E}(D \mid \mathcal{T}_2)\mu(\mathcal{T}_2)^2 + 4h\mu(\mathcal{T}_1)\mu(\mathcal{T}_2).
\end{aligned}$$

We have essentially used the fact that all the leaves are at distance h from r, thus the distance between any $l \in \mathcal{T}_1$ and $l' \in \mathcal{T}_2$ is $d(l,l') = d(l,r) + d(l',r) = 2h$. □

Recursive application of Proposition 1 yields a linear algorithm that computes $\mathbb{E}(D)$ for a given probability distribution μ on the leaves of $\mathcal{T}$. It is presented as Algorithm 1. It computes the values of $\mathbb{E}(D \mid \mathcal{T}_v)$ in one depth-first traversal of $\mathcal{T}$.

Algorithm 1 Computing quadratic entropy for a given probability distribution.

Procedure `compute_`$\varepsilon(v,d)$
Parameter v**:** vertex for which we are computing $\mathbb{E}(D|\mathcal{T}_v)$.
Parameter d**:** distance from v to the leaves of $\mathcal{T}_v$.

if v is a leaf **then**
  set $\varepsilon(v) = 0$.
  let $\mu(v)$ be the assigned probability $\mathbb{P}(l = v)$.
else
  let $c_1, \ldots, c_t$ be the children of v.
  `compute_`$\varepsilon(c_1, d - d(vc_1))$.
  set $\varepsilon(v) = \varepsilon(c_1)$.
  set $\mu(v) = \mu(c_1)$
  for $i = 2$ to t **do**
    `compute_`$\varepsilon(c_i, d - d(vc_i))$.
    set $\mu(v) = \mu(v) + \mu(c_i)$.
    set $p = \mu(c_i)/\mu(v)$.
    set $\varepsilon(v) = \varepsilon(c_i)p^2 + \varepsilon(v)(1-p)^2 + 4dp(1-p)$.
  end for
end if

For the proof of correctness in this and the following sections, we introduce some notation. Let $v \in V(\mathcal{T})$ be some vertex of $\mathcal{T}$, and let $c_1, \ldots, c_t$ be its children. In this context, let $\mathcal{T}_i$ be the tree, rooted at v, obtained recursively as $\mathcal{T}_1 := \mathcal{T}_{c_1} \cup vc_1$ and $\mathcal{T}_i = \mathcal{T}_{i-1} + (\mathcal{T}_{c_i} \cup vc_i)$ for $i \geq 2$, where $\mathcal{T}_{c_i} \cup vc_i$ is the tree obtained from $\mathcal{T}_{c_i}$ by adding the edge vc_i. Note that $\mathbb{E}(D \mid \mathcal{T}_{c_i} \cup vc_i)$ and $\mu(\mathcal{T}_{c_i} \cup vc_i)$ are the same as $\mathbb{E}(D \mid \mathcal{T}_{c_i})$ and $\mu(\mathcal{T}_{c_i})$, respectively.

Theorem 2. *For $v \in V(\mathcal{T})$, the value $\varepsilon(v)$ computed by Algorithm 1 equals $\mathbb{E}(D \mid \mathcal{T}_v)$. In particular, $\varepsilon(r)$ is the quadratic entropy of $\mathcal{T}$ for a given probability distribution μ.*

Proof. In addition to the statement of the theorem, we claim that $\mu(v) = \mathbb{P}(l \in \mathcal{T}_v)$. We prove these claims by induction on the number of vertices in $\mathcal{T}_v$. If there are only two vertices $r = u$ and v, which is the leaf, then $\mu(v) = 1$ (line 3), $\varepsilon(v) = 0$ (line 2), $\mu_r(u) = 1$ (line 8), and $\varepsilon(u) = 0$ (lines 6 and 7), which is correct.

Let there be at least three vertices in $\mathcal{T}_v$. Correct results are computed for the children c_i of v_i in lines 6 and 10 by induction. If there is only one child, then $\mu(c_i) = \mu(\mathcal{T}_v)$ by induction, and lines 8 and 7 establish correctness of $\mu(v)$ and $\varepsilon(v)$.

If there are more children, then line 11 computes the probability $\mu(\mathcal{T}_i)$ and line 12 computes $\mathbb{P}(l \in \mathcal{T}_{c_i} \mid l \in \mathcal{T}_i)$ by conditional probabilities. Line 13 computes $\mathbb{E}(D \mid \mathcal{T}_i)$ by Proposition 1. Then $\varepsilon(v) = \mathbb{E}(D \mid \mathcal{T}_v)$ and $\mu(v) = \mu(\mathcal{T}_v)$ after the execution of the `for` loop, since $\mathcal{T}_t = \mathcal{T}_v$. The theorem follows. □

4. Quadratic entropy versus weighted Wiener index

In this section, we show that there is a close connection between the quadratic entropy and one of the central concepts studied in chemical graph theory. Extending the methods from this field of research we give an alternative algorithm (Corollary 4) for computing the quadratic entropy of $\mathcal{T}$. This algorithm avoids computing the distances between the leaves.

In mathematical chemistry, numerous graph invariants are used to analyze and predict physical and chemical properties of chemical compounds. When such invariants are computed on chemical graphs, they are traditionally called *topological indices*. Among topological indices, the Wiener index is the oldest [35] and one of the most thoroughly studied indices, see, e.g. the surveys [31,36]. Let $G = (V(G), E(G))$ be a connected graph. Then the *Wiener index* $W(G)$ of G is defined as the sum of the shortest path distances between all unordered pairs of vertices:

$$W(G) = \sum_{\{u,v\} \subseteq V(G)} d(u, v) = \frac{1}{2} \sum_{u,v \in V(G)} d(u, v).$$

This classical definition was extended in [33] to weighted graphs (G, f), where $f : V(G) \to \mathbb{R}$ is a vertex weighting function, in the following way:

$$W(G, f) = \frac{1}{2} \sum_{u,v \in V(G)} f(u) f(v) d(u, v).$$

Note that in the definitions of the (weighted) Wiener index it is assumed that all the edges have unit length, that is, the w-values on its edges are all 1.

In order to design a linear algorithm for computing the Wiener index of an important class of chemical graphs—benzenoid systems—it was observed in [37] that the Wiener index of a weighted tree can be computed in linear time. (This result in also implicit in [34].) We now show that the approach can be extended to weighted trees with edges of arbitrary length. The weighted Wiener index $W(G, f)$ is defined as before, except that now $d(u, v)$ is the sum of the w-values on a shortest u, v-path.

Let (T, f) be a weighted tree and let uv be an edge of T. Then $T - uv$ consists of connected components, say T^u and T^v, where $u \in T^u$ and $v \in T^v$. Let $f(T^u) = \sum_{x \in T^u} f(x)$ and $f(T^v) = \sum_{x \in T^v} f(x)$.

Proposition 3. *Let (T, f) be a weighted tree with w-values on its edges. Then*

$$W(T, f) = \sum_{uv \in E(T)} f(T^u) f(T^v) w(u, v).$$

Proof. Let uv be an arbitrary edge of T and let $x, y \in V(T)$. Then e lies on the u, v-path if and only if one of x, y belongs to T^u and the other to T^v. Suppose that this is the case and let $x = x_1, \ldots, x_j = u, x_{j+1} = v, \ldots, x_k = y$ be the x, y-path in T. Then

$$f(u) f(v) d(u, v) = f(u) f(v) \sum_{i=1}^{k-1} w(x_i, x_{i+1}).$$

Hence the contribution of the edge uv to $W(G, f)$ with respect to the unordered pair x, y is $f(x) f(y) w(u, v)$. Since this holds for all pairs of vertices from T^u and T^v, the result follows. □

Corollary 4. *Let (T, r, w) be an evolutionary tree with probability distribution μ on its leaves. Then*

$$\mathbb{E}(D) = 2 \cdot \sum_{e_v \in E(T)} w(e_v)\mu(T_v)(1 - \mu(T_v)).$$

Proof. Define $f : V(T) \to \mathbb{R}$ with

$$f(u) = \begin{cases} \mu(u); & u \text{ is a leaf of } T, \\ 0; & \text{otherwise.} \end{cases}$$

Then note that $\mathbb{E}(D) = 2W(T, f)$ and apply Proposition 3. □

Special cases of vertex-weighted Wiener indices (and their Wiener polynomial) were recently treated in [38,32], where the assigned weights are vertex degrees. The general case, in which both vertices and edges are weighted, has been to the best of our knowledge treated earlier only by Zmazek and Žerovnik [39]. They give a linear algorithm for cactus graphs, the graphs whose blocks are cycles and edges. Hence their (rather involved) algorithm can be considered as an extension of the algorithm that flows from Corollary 4.

5. Maximizing quadratic entropy

In this section, we present an algorithm that computes the maximum value of quadratic entropy over all possible probability distributions on the leaves of $\mathcal{T}$, together with the probability distribution μ on the leaves of $\mathcal{T}$ that achieves the maximum value. In terms of Wiener index, this problem finds the weighting function on the set of leaves of $\mathcal{T}$, such that the resulting weighted Wiener index is maximum. This weighting is Pavoine's originality score [22] from conservation biology. To our knowledge, this problem has not been studied earlier in the Wiener index framework.

Proposition 5. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two trees of height h with probability distributions μ_1, μ_2 and expected distances $\mathbb{E}(D_1)$, $\mathbb{E}(D_2)$. Further, let $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2$ be their join. Then the distribution μ, defined with*

$$\mu(l) = \begin{cases} \dfrac{2h - \mathbb{E}(D_2)}{4h - \mathbb{E}(D_1) - \mathbb{E}(D_2)}\mu_1(l); & l \in \mathcal{T}_1 \\ \dfrac{2h - \mathbb{E}(D_1)}{4h - \mathbb{E}(D_1) - \mathbb{E}(D_2)}\mu_2(l); & l \in \mathcal{T}_2 \end{cases} \tag{5.1}$$

maximizes $\mathbb{E}(D)$ whenever μ_1 and μ_2 maximize $\mathbb{E}(D_1)$ and $\mathbb{E}(D_2)$.

Proof. First note that $\mu(\mathcal{T}) = 1$ and that $\mu(\mathcal{T}_i)$ is obtained by scaling μ_i, $i = 1, 2$, therefore $\mathbb{E}(D \mid \mathcal{T}_i) = \mathbb{E}(D_i)$. By $\mu(\mathcal{T}_1) + \mu(\mathcal{T}_2) = 1$ and Proposition 1, we have

$$\mathbb{E}(D) = \mathbb{E}(D \mid \mathcal{T}_1)\mu(\mathcal{T}_1)^2 + \mathbb{E}(D \mid \mathcal{T}_2)(1 - \mu(\mathcal{T}_1))^2 + 4h\mu(\mathcal{T}_1)(1 - \mu(\mathcal{T}_1)).$$

This expression involves the constant h and three variables, $\mu(\mathcal{T}_1)$, $\mathbb{E}(D \mid \mathcal{T}_1)$ and $\mathbb{E}(D \mid \mathcal{T}_2)$, which are not independent. We optimize $\mathbb{E}(D)$ under the assumption of their independence, which we justify later.

For fixed $\mathbb{E}(D \mid \mathcal{T}_1)$, $\mathbb{E}(D \mid \mathcal{T}_2)$, and variable $\mu(\mathcal{T}_1)$, $\mathbb{E}(D)$ is maximized in the apex of the parabola, thus

$$\mu(\mathcal{T}_1) = \frac{2h - \mathbb{E}(D \mid \mathcal{T}_2)}{4h - \mathbb{E}(D \mid \mathcal{T}_1) - \mathbb{E}(D \mid \mathcal{T}_2)}$$

implying

$$\mu(\mathcal{T}_2) = \frac{2h - \mathbb{E}(D \mid \mathcal{T}_1)}{4h - \mathbb{E}(D \mid \mathcal{T}_1) - \mathbb{E}(D \mid \mathcal{T}_2)}.$$

Using these values, we obtain the maximum

$$\mathbb{E}(D) = \frac{4h^2 - \mathbb{E}(D \mid \mathcal{T}_1)\mathbb{E}(D \mid \mathcal{T}_2)}{4h - \mathbb{E}(D \mid \mathcal{T}_1) - \mathbb{E}(D \mid \mathcal{T}_2)}.$$

The partial derivative in variables $\mathbb{E}(D \mid \mathcal{T}_1)$ and $\mathbb{E}(D \mid \mathcal{T}_2)$ is everywhere nonnegative, thus $\mathbb{E}(D)$ will be maximized when both $\mathbb{E}(D \mid \mathcal{T}_1)$ and $\mathbb{E}(D \mid \mathcal{T}_2)$ will be largest. By assumption, this is achieved if μ restricted to $\mathcal{T}_1$ equals μ_1 and μ restricted to $\mathcal{T}_2$ equals μ_2.

The distribution μ described by formula (5.1) satisfies all three conditions: $\mu(\mathcal{T}_1)$ and $\mu(\mathcal{T}_2)$ have the desired value and μ restricted to $\mathcal{T}_i$ is after normalization equal to μ_i, $i = 1, 2$. Since the distribution μ satisfies the optimality conditions for the optimum without dependence of the three variables, it achieves the optimum in the restricted case and therefore maximizes $\mathbb{E}(D)$. □

Algorithm 2 First pass for computing relative subtree probabilities.

Procedure `maximize_`$\varepsilon(v,d)$
Parameter v**:** vertex for which computation is done.
Parameter d**:** distance from v to the leaves of $\mathcal{T}_v$.
1: set $\mu_r(v) = 1$.
2: **if** v is a leaf **then**
3: set $\varepsilon(v) = 0$.
4: **else**
5: let $c_1, \ldots, c_t$ be the children of v.
6: `maximize_`$\varepsilon(c_1, d - d(vc_1))$.
7: set $\varepsilon(v) = \varepsilon(c_1)$.
8: **for** $i = 2$ to t **do**
9: `maximize_`$\varepsilon(c_i, d - d(vc_i))$.
10: set $\mu_r(c_i) = \frac{2d-\varepsilon(v)}{4d-\varepsilon(c_i)-\varepsilon(v)}$.
11: set $\varepsilon(v) = \varepsilon(c_i)\mu_r(c_i)^2 + \varepsilon(v)(1-\mu_r(c_i))^2 + 4d\mu_r(c_i)(1-\mu_r(c_i))$.
12: **end for**
13: set $x = 1 - \mu_r(c_t)$.
14: **for** $i = t - 1$ downto 1 **do**
15: set $y = 1 - \mu_r(c_i)$.
16: set $\mu_r(c_i) = \mu_r(c_i)x$.
17: set $x = xy$.
18: **end for**
19: **end if**

We use Propositions 1 and 5 recursively in Algorithm 4, which computes μ in two passes of depth-first traversing of $\mathcal{T}$. In the first pass, Algorithm 2, we compute $\varepsilon(v) = \mathbb{E}(D \mid \mathcal{T}_v)$ for every vertex v of $\mathcal{T}$ and $\mu_r(v) = \mathbb{P}(l \in \mathcal{T}_v \mid l \in \mathcal{T}_u)$, i.e. the probability for a leaf selected from $\mathcal{T}_u$ to lie in $\mathcal{T}_v$, where u is the parent of v. In the second pass, Algorithm 3, we compute absolute probabilities $\mu(v) = \mu(\mathcal{T}_v)$ for a leaf to be selected in $\mathcal{T}_v$. A child of some vertex tree is considered at most twice in Algorithm 2 and once in Algorithm 3, thus the algorithm is linear in the number of vertices of the tree. The following Theorem establishes its correctness.

Algorithm 3 Second pass for computing absolute subtree probabilities.

Procedure `compute_`$\mu(v,\tau)$
Parameter v**:** the vertex for which the computation is done.
Parameter τ**:** the value $\mu(\mathcal{T}_u)$ for u the parent of v.
1: set $\mu(v) = \mu_r(v)\tau$.
2: **for each** child c of v **do**
3: `compute_`$\mu(c, \mu(v))$.
4: **end for**

Algorithm 4 Recursive calls maximizing the quadratic entropy.

1: `maximize_`ε`_`$\mu_r(r, h)$.
2: `compute_`$\mu(r,1)$.

Theorem 6. *The probability distribution μ computed by Algorithm 4 maximizes the quadratic entropy of the evolutionary tree $\mathcal{T}$. The value $\varepsilon(r)$ stores the maximum quadratic entropy.*

Proof. First we prove by induction on the number of vertices in $\mathcal{T}_v$ that Algorithm 2 correctly computes $\varepsilon(v) = \mathbb{E}(D \mid T_v)$ and $\mu_r(v) = \mathbb{P}(l \in \mathcal{T}_v \mid l \in \mathcal{T}_u)$, u being the parent of v. If there are only two vertices $r = u$ and v, which is the leaf, then $\mu_r(v) = 1$ (line 1), $\varepsilon(v) = 0$ (line 3), $\mu_r(u) = 1$ (line 1), and $\varepsilon(u) = \varepsilon(v) = 0$ (lines 6 and 7), which is correct.

Let there be at least three vertices in $\mathcal{T}_v$. By induction, each call in lines 6 and 9 computes correct information for $\mathcal{T}_{c_i}$, where c_i is some child of v. It is easy to see that the same values apply to the non-root vertices in the tree $\mathcal{T}_{c_i} \cup vc_i$ rooted at v. If there is only one child, then lines 1 and 7 assure correctness of $\mu(v)$ and $\varepsilon(v)$.

If there are more children, then line 10 computes the correct value $\mu_r(c_i)$ relative to the tree $\mathcal{T}_i$ by Proposition 5, and line 11 correctly computes $\mathbb{E}(D \mid \mathcal{T}_i)$ by Proposition 1. Thus $\mu_r(c_t)$ and $\varepsilon(v)$ are computed correctly in the first loop. For other children of v, at each join evaluated in lines 10 and 11, we would by Proposition 5 need to update $\mu_r(c_j)$, $1 \leq j < i$, by a factor of $1 - \mu_r(c_i)$, where the latter is the value computed in line 10. This is done in line 17: as the value y accumulates the updating factor, only one visit to each child is necessary for update. The correctness of Algorithm 2 follows.

By conditional probabilities,

$$\mathbb{P}(l \in \mathcal{T}_v) = \mathbb{P}(l \in \mathcal{T}_v \wedge l \in \mathcal{T}_u) = \mathbb{P}(l \in \mathcal{T}_u)\mathbb{P}(l \in \mathcal{T}_v \mid l \in \mathcal{T}_u).$$

Thus, $\mu(v) = \mu(u)\mu_r(v)$, which via line 1 of Algorithm 3 establishes correctness of Algorithm 3. We correctly set $\mu(r) = 1$ in line 2 of Algorithm 4. Since $\mu(v) = \mu(\mathcal{T}_v)$ for any leaf v of $\mathcal{T}$, we conclude the proof. □

6. Concluding remarks

We present several new insights into quadratic entropy of ultrametric evolutionary (but not necessarily binary) trees, drawn from computational chemistry and graph theory. First, we propose a recursive decomposition of evolutionary trees and derive a formula to express quadratic entropy of the whole tree as a function of quadratic entropies of the subtrees in the decomposition (Proposition 1). Quadratic entropy for trees with such defined weights (e.g. abundances) can be used as a community biodiversity index that incorporates evolutionary history and community structure [25]. We use Proposition 1 to design an efficient linear time algorithm (Algorithm 1) that computes the quadratic entropy of such trees and can handle large communities (cf. [40]). Second, we observe that quadratic entropy of an evolutionary tree is a variant of weighted Wiener index, a general graph invariant widely used in mathematical chemistry. This link may establish an exchange of ideas between the areas, as demonstrated by the statements of Section 4, where we expose some theoretical properties of quadratic entropy—Wiener index. This relationship can be utilized to provide a linear time algorithm for computing quadratic entropy on arbitrary edge- and vertex-weighted tree (not necessarily ultrametric nor binary). Finally, maximizing the quadratic entropy offers a novel pairwise originality metric [22,23] whose properties still remain relatively unexplored (but see [26]). We derive a linear time algorithm for its maximization on ultrametric evolutionary trees (not necessarily binary), Algorithm 4, that supersedes existing algorithms, and may help in exploration of this quantity. In this algorithm, the fact that all leaves of a given subtree have the same distance to this vertex plays an essential role; would these distances have been different, an equivalent of Proposition 1 would have to consider the structure of the two trees in the join, not just their respective entropies. Similarly, the coefficient for updating the weights of the optimal solutions of the subtrees would be different for each leaf. These observations imply that any algorithm for maximizing quadratic entropy on non-ultrametric trees resulting from our approach would be significantly more complex than ours. In particular, it would not be linear, and therefore we do not pursue this direction.

Acknowledgments

The authors appreciate useful comments of two anonymous referees that provided some polishing of the original presentation.

The first author was supported in part by the Ministry of Higher Education, Science and Technology of Slovenia, Research Project L1-5014 and Research Program P1-0297. The third author was supported in part by the Ministry of Science of Slovenia under the grant P1-0297. The fifth author was supported by a fellowship from the Wissenschaftskolleg zu Berlin, Wallotstrasse 19, 14193 Berlin, Germany, and NSERC Canada.

References

[1] C. Semple, M. Steel, Phylogenetics, Oxford University Press, Oxford, 2003.
[2] M. Baroni, C. Semple, M. Steel, A framework for representing reticulate evolution, Ann. Comb. 8 (2004) 391–408.
[3] J. Felsenstein, Inferring Phylogenies, Sinauer Associates, Sunderland, Massachusetts, 2004.
[4] G.U. Yule, A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, Philos. Trans. R. Soc. Lond. Ser. B—Biological Sciences 213 (1924) 21–87.
[5] D.E. Rosen, Vicariant patterns and historical explanation in biogeography, Systematic Zoology 27 (1978) 159–188.
[6] A.Ø. Mooers, L.J. Harmon, M.G.B. Blum, D.H.J. Wong, S.B. Heard, Some models of phylogenetic tree shape, in: O. Gascuel, M. Steel (Eds.), Reconstructing Evolution: New Mathematical and Computational Advances, Oxford University Press, Oxford, 2007, pp. 149–170.
[7] M. Steel, A. McKenzie, Properties of phylogenetic trees generated by Yule-type speciation models, Math. Biosci. 170 (2001) 91–112.
[8] A.W.F. Edwards, Estimation of branch points of a branching diffusions process, (with discussion) J. R. Stat. Soc. B 32 (1970) 155–174.
[9] Z. Yang, B. Rannala, Branch-length prior influences Bayesian posterior probability of phylogeny, Systematic Biology 54 (2005) 455–470.
[10] C. Semple, M. Steel, A supertree method for rooted trees, Discrete Appl. Math. 105 (2000) 147–158.
[11] C. Semple, Reconstructing minimal rooted trees, Discrete Appl. Math. 127 (2003) 489–503.
[12] J. Hey, Using phylogenetic trees to study speciation and extinction, Evolution 46 (1992) 627–640.
[13] S. Nee, R.M. May, P.H. Harvey, The reconstructed evolutionary process, Philos. Trans. R. Soc. Ser. B–Biological Sciences 344 (1994) 305–311.
[14] D.L. Rabosky, Likelihood methods for detecting temporal shifts in diversification rates, Evolution 60 (2006) 1152–1164.
[15] D.P. Faith, Conservation evaluation and phylogenetic diversity, Biological Conservation 61 (1992) 1–10.
[16] K. Hartmann, M. Steel, Maximising phylogenetic diversity in biodiversity conservation: greedy solutions to the Noah's ark problem, Systematic Biology 55 (2006) 644–651.
[17] M.L. Weitzman, The Noah's ark problem, Econometrica 66 (1998) 1279–1298.
[18] C.-J. Haake, A. Kashiwada, F.E. Su, The Shapley value of phylogenetic trees, IMW Working Paper 363, (2005).
[19] K. Hartmann, M. Steel, Phylogenetic diversity: from combinatorics to ecology, in: O. Gascuel, M. Steel (Eds.), Reconstructing Evolution: New Mathematical and Computational Approachese, Oxford University Press, Oxford, 2007.
[20] N.J.B. Isaac, S.T. Truvery, B. Colen, C. Waterman, J.E.M. Baillie, Mammals on the edge: conservation priorities based on threat and phylogeny, PLoS One 2 (3) (2007) e296.
[21] R.M. May, Taxonomy as destiny, Nature 347 (1990) 129–130.
[22] S. Pavoine, S. Ollier, A.B. Dufour, Is the originality of a species measurable? Ecology Letters 8 (2005) 579–586.

[23] D.W. Redding, Incorporating genetic distinctness and reserve occupancy in a conservation priorisation approach, Masters Thesis, University Of East Anglia, Norwich, 2003.
[24] D.W. Redding, A.Ø. Mooers, Incorporating evolutionary measures into conservation prioritization, Conservation Biology 20 (2006) 1670–1678.
[25] C.R. Rao, Diversity and dissimilarity coefficients: a unified approach, J. Theoret. Pop. Bio. 21 (1982) 24–43.
[26] S. Pavoine, S. Ollier, D. Pontier, Measuring diversity from dissimilarities with Rao's quadratic entropy: are any dissimilarities suitable? J. Theoret. Pop. Bio. 67 (2005) 231–239.
[27] S. Champely, S. Pavoine, divcmax: Maximal value of Rao's diversity coefficient also called quadratic entropy, in: D. Chessel, A.B. Dufour, S. Dray (Eds.), The ade4 Package, Analysis of Environmental Data: Exploratory and Euclidean Methods in Environmental Sciences, Université Claude Bernard Lyon 1, 2006, http://microarrays.unife.it/CRAN/doc/packages/ade4.pdf.
[28] Analyses des Données Ecologiques: méthodes Exploratoires et Euclidiennes en sciences de l'Environnement, Centre National De La Recherche Scientifique et Université Claude Bernard Lyon 1, Lyon, 2007. http://pbil.univ-lyon1.fr/ADE-4.
[29] The R Project for Statistical Computing, Department of Statistics and Mathematics, WU Wien, 2007. http://www.r-project.org/.
[30] I. Martyn, iain.martyn@mail.mcgill.ca.
[31] A.A. Dobrynin, R. Entringer, I. Gutman, Wiener index of trees: theory and applications, J. Acta Appl. Math. 66 (2001) 211–249.
[32] T. Došlić, Vertex-weighted Wiener polynomials for composite graphs, Ars Math. Contemp. 1 (2008) 66–80.
[33] S. Klavžar, I. Gutman, Wiener number of vertex-weighted graphs and a chemical application, Discrete Appl. Math. 80 (1997) 73–81.
[34] B. Mohar, T. Pisanski, How to compute the Wiener index of a graph, J. Math. Chem. 2 (1988) 267–277.
[35] H. Wiener, Structural determination of paraffin boiling points, J. Amer. Chem. Soc. 69 (1947) 17–20.
[36] A.A. Dobrynin, I. Gutman, S. Klavžar, P. Žigert, Wiener index of hexagonal systems, J. Acta Appl. Math. 72 (2002) 247–294.
[37] V. Chepoi, S. Klavžar, The Wiener index and the Szeged index of benzenoid systems in linear time, J. Chem. Inf. Comput. Sci. 37 (1997) 752–755.
[38] D.J. Klein, T. Došlić, D. Bonchev, Vertex-weightings for distance moments and thorny graphs, Discrete Appl. Math. 155 (2007) 2294–2302.
[39] B. Zmazek, J. Žerovnik, Computing the weighted Wiener and Szeged number on weighted cactus graphs in linear time, Croat. Chem. Acta 76 (2003) 137–143.
[40] M. Vellend, W. Cornwell, K. Magnuson-Ford, A.Ø. Mooers, Measuring phylogenetic biodiversity, in: A. Magurran, B. McGill (Eds.), Biological Diversity: Frontiers in Measurement and Assessment, Oxford University Press, New York, 2011.