

Tíz kiemelt hivatkozás - idézetekkel

Erdős Péter

K. Atteson: The performance of neighbor-joining methods of phylogenetic reconstruction, *Algorithmica* 25 (1999), 251–278. Hivatkozta:[20]

P. 256.: The later result also holds for the UNJ and BIONJ methods of Gasquel [G2], [G3] which are modifications of NJ. Note that methods described in [ESSW] and the Buneman tree method [Bu] are also known to have this property but known algorithms implementing these methods have higher computational complexity than some of the neighbor-joining methods. A method which finds the closest additive distance matrix to the input distance matrix under the l_∞ norm would have l_∞ at least $\frac{1}{4}$ (see [ESSW]). However, this problem is NP-hard to approximate within a factor of $\frac{9}{8}$ [ABF+]. A 3-approximation to this problem is known [ABF+] which has l_∞ radius between $\frac{1}{8}$ and $\frac{1}{6}$ (see [ESSW]). Motivated by Lemma 3, we now give a name to a distance matrix which is near enough to a weighted binary tree so that it can be guaranteed to be correctly reconstructed by a method with optimal l_∞ radius:

The concept of nearly additive distance matrices was introduced in [ESSW] (without the name).

D. Bryant: Extending tree models to split networks, Chapter 17, in *Algebraic Statistics for Computational Biology* (Ed. L. Pachter and B. Sturmfels) Cambridge Univ. Press (2005), 331–346. Hivatkozta:[9, 8]

P. 332.: **17.6 A Fourier calculus for splits networks**
[Székely *et al.* 1993] describe a *Fourier calculus on evolutionary trees* that generalizes the Hadamard transform [Hendy and Penny, 1989, Steelet *al.*, 1992]. Using their approach, we can take the observed character frequencies, apply a transformation, and obtain a vector of values from which we can read off the support for different splits. They show that if the observed character frequencies correspond exactly to the character probabilities determined by some phylogenetic tree then the split supports will correspond exactly to the splits and branch length in the phylogenetic tree. Conversely, the inverse transformation gives a single formula for the character probabilities in any tree. This theory generalizes seamlessly from trees to splits networks - in fact so seamlessly that the proofs of [Székely *et al.*, 1993] requires almost no modification to establish the general case.

M. Csűrös - K-Y. Ming: Recovering evolutionary trees through Harmonic Greedy Triplets. *SODA '99 - Tenth Annual ACM-SIAM Symposium on Discrete Algorithms*, (1999), 1–12. Hivatkozta:[17]

P. 261.: At present, the Short Quartet Method (SQM) by Erdős *et al.* is the only other known algorithm with comparable theoretical and experimental performance. ... The SQM algorithm was originally analyzed for the Cavander-Farris model [3], which is the special binary-character case of the Jukes-Cantor model. The analysis techniques presented in this paper can be used to extend the theoretical analysis of SQM algorithm to the Jukes-Cantor model. The theoretical and experimental performance of both HGT and SQM algorithms have demonstrated to be superior to such other distance based algorithms as that of Farrach and Kannan [7] and the widely used Neighbor Joining [17] algorithm.

E. Dahlhaus - D.S. Johnson - C.H. Papadimitriou - P.D. Seymour - M. Yannakakis: The complexity of multiterminal cuts, *SIAM J. Computing* **23** (1994), 864–894. Hivatkozsa:[5, 6] (extended abstract: 24th ACM STOC, (1992), 241–251. Hivatkozsa:[5, 12]).

P. 892.: More recently, Erdős and Székely in [5], [6] proposed the following generalization of multiterminal cut. Suppose you are given a graph $G = (V, E)$ with weighted edges, and a partial k -coloring of the vertices, that is, a subset and a function $f : V' \rightarrow \{1, 2, \dots, k\}$. Can f be extended to a total function such that the total weight of the edges that have different colored endpoints is minimized? The k -terminal cut problem is the special case where $|V'| = k$ and F is $1 - 1$, that is, each color is initially assigned to precisely one vertex. It is easy to see that for general graphs, this problem is in fact equivalent to multiterminal cut. ...

Nevertheless, in the case of trees, the dynamic programming algorithm for multiterminal cut mentioned in the Introduction extends in a natural way to the colored multiterminal cut problem, yielding an $O(nk)$ algorithm, as Erdős and Székely observe. This, in turn, implies that if G is such that deleting all the terminals renders it acyclic, then multiterminal cut can itself still be solved in $O(nk)$ time. (Simply split each terminal s_i into $\text{degree}(s_i)$ separate vertices, one for each edge incident on s_i , assign color to all the derived vertices, and apply the above mentioned algorithm for colored multiterminal cut on trees to the resulting graph [6]).

Duffus - B. Sands: Minimum sized fibres in distributive lattices, *Austr. J. Math* **70** (2001), 337–350. Hivatkozsa:[13, 15]

P. 339.: At this point, we can prove somewhat less than this. Our result, Theorem 2, depends on what Ahlswede, Erdős and Graham [1] call the 'splitting property' for maximal antichains. Call an ordered set X *dense* if every proper nonempty open interval $(a, b) = \{x \in X \mid a < x < b\}$ of X contains at least two elements. Say that a maximal antichain A of X has the *splitting property* if A can be partitioned into two subsets B and C so that $X = B \uparrow \cup C \downarrow$. X has the *splitting property* if all of its maximal antichains do. The splitting property for infinite antichains is studied in ([7], [8]).

The important results for us from [1] are that *Boolean lattices are dense* and *every dense ordered set has the splitting property*. ... We will show that the splitting property is nicely tied to the monotonicity of f , in all three meanings contained in Conjecture 2.

K. Engel: *Sperner Theory, Encyclopedia of Mathematics and Its Applications, Vol. 65 Cambridge University Press, 1997.* Hivatkozsa:[1, 2, 3, 7, 14] A könyv 3. Fejezete (84-115 oldalak) részletesen tárgyalja a halmazrendszerek konvex burkának elméletét.

P. 162.: For a subset (family) F of P and an automorphism ϕ , let $\phi(F) := \{\phi(p) : p \in F\}$. Given an automorphism group G of P , we say that a class $\mathfrak{A}$ of families in P is G -invariant if $\phi(F) \in \mathfrak{A}$ for all $F \in \mathfrak{A}$ The following theorem is (in a different formulation) due to Erdős, Faigle, and Kern [173].

Theorem 4.5.7. Suppose that there exists a rank-transitive automorphism group of P . Let $C = (p_0 < p_1 < \dots < p_n)$ be a fixed maximal chain in P , let Q be the filter generated by p_0 and let R be a system of representatives of the left cosets of G relative to G_{p_0} . Let $w : P \rightarrow \mathbb{R}_+$ be defined by

$$w(p) := \frac{W_i(P)}{W_0(P)} \frac{1}{W_i(Q)} \quad \text{if } p \in N_i(P),$$

and let $\mathfrak{A}$ be a G -invariant class of families. If for all $F \in \mathfrak{A}$

$$\sum_{\rho \in R} w(\rho(C) \cap F) \leq 1,$$

then for all $F \in \mathfrak{A}$

$$\sum_{i=0}^n \frac{f_i}{W_i(Q)} \leq 1.$$

J. Felsenstein: *Inferring phylogenies*, Sinauer Associates Inc. Sunderland CT, (2003), 580pp p.173,182 Hivatkozsa:[8, 11, 16, 18, 19]

P. 172.: A puzzling formula

Erdős et al. (1997a) give two versions of these bounds.

$$k > \frac{c \log n}{f^2(1 - 2g)^{2\text{diam}(T)}} \quad (11.21)$$

The result is surprising because it seems to imply that we need only have a number of characters proportional to the logarithm of the number of species,.....

P. 182-3.: Short quartet methods

The noise can easily arise if some of the species are rather distant from each other. This is also a serious problem with distance matrix methods such as neighbor-joining and those using the Fitch-Margolis criterion.

To correct this, Erdős et al. (1997a,1997b,1999) have put forward the *short quartet method*. This reconstruct a tree from quartets that do not involve any of the large distances. This method uses a threshold value of distance and accept only those quartets that do not have any of the distances between their members greater than this threshold.... Inferring trees from these „short” quartets, they then combine them to make an estimate of the overall tree. The method of combination used is complete compatibility of the quartets.....

P. 286.: Extensions of Hadamard methods

Steel et al. (1993) have shown that if the distribution of rates r among sites is $f(r)$, then we must replace the logarithm in equations 17.10 and 17.12 by the inverse of the moment-generating function of $f(r)$. This is

$$M^{-1}(r) = \int_0^\infty \ln(\lambda r) f(\lambda) d\lambda \quad (17.16)$$

If this function can be evaluated (which it can for distributions like as gamma distributions) then it used instead of the logarithm allows us to have a Hadamard conjugation that works for the model with varying rates among sites.

E. Mossel - S. Roch: *Learning nonsingular phylogenies and hidden Markov models*, *Ann. Appl. Probability* **16** (2) (2006), 583–614. Hivatkozsa:[19, 20]

P. 588.: Reconstructing the topology has been a major task in phylogeny. It follows from [10, 11] that the topology can be recovered with high probability using a polynomial number of samples. Here is one formulation from [26].

THEOREM 3. Let $\beta > 0, \kappa_\beta > 0$ and suppose that $\mathbf{M}_n$ consists of all matrices P satisfying $\beta < |\det P| < 1 - n^{-\kappa_\beta}$. For all $\kappa_T > 2$, the topology of $T \in (\mathbf{T}_3(n) \otimes \mathbf{M}_n, n^{-\kappa_\pi})$ can be recovered in polynomial time using $n^{O(1/\beta + \kappa_\beta + \kappa_T + \kappa_\pi)}$ samples with probability at least $1 - n^{2-\kappa_T}$.

We will also need a stronger result that applies only to hidden Markov models. The proof, which is sketched in the Appendix, is quite similar to the proofs in [10, 11].

P. 610.: A crucial observation in [10, 11] is that, to obtain good estimates of distances with a polynomial number of samples, one has to consider only pairs of leaves at a „short” distance. We note $\hat{\Psi}_{ab}$ the estimate of Ψ_{ab} . For $\Delta > 0$, define

$$S_\Delta = \{(a, b) \in \mathcal{L} \times \mathcal{L} : \hat{\Psi}_{ab} > 2\Delta\}.$$

Let $\Delta = -\ln[6n^{-\zeta}]$. Then it follows from [11], Proof of Theorem 14, that, for any $e, p > 0$, there exists an $s > 0$ large enough so that, using n^s samples, with probability at least $1 - n^{-p}$, one has, for all (a, b) in $S_{2\Delta}$,

$$|\hat{\Psi}_{ab} - \Psi_{ab}| < -\ln[1 - n^{-e}] \leq n^{-e},$$

and $S_{2\Delta}$ contains all pairs of leaves with $\Psi_{ab} \leq 2\Delta$, but no pair with $\Psi_{ab} \geq 6$

By [10], Lemma 5, this is guaranteed to return the right topology if, for all $a', b' \in q$,

$$|\hat{\Psi}_{a'b'} - \Psi_{a'b'}| < \frac{x}{2},$$

where x is the length (in the log-det distance) of the internal edge in the subtree induced by q .

M. Pouly: Minimizing Communication Costs of Distributed Local Computation., in ECAI'2006, Workshop 26 (ed. A. Darwiche *et al.*), (2006), 19–24. Hivatkozsa:[12]

P. 23.:

4.1 Analysis of PDP

[1] originally initiated the study of the multiway cut problem whose close relation to PDP is the topic of this section. For our purposes, we stress the more general and illustrative definition of multiway cut given in [2]: ...

This definition includes the so-called color-independent version of a weight function, which has also been used in [1]. A more general form is proposed by [2] and defined as $w : E \times C \times C \rightarrow \mathbb{N}$. In this case, the weight function is called color-dependent and the number $w(i, j, p, q)$ specifies the weight of the edge $(i, j) \in E$, if $\bar{v}(i) = p$ and $\bar{v}(j) = q$. Clearly, color-independence is reached, if for any $(i, j) \in E, p_1 \neq q_1$ and $p_2 \neq q_2$, we have $w(i, j, p_1, q_1) = w(i, j, p_2, q_2)$. Finally, if $w(i, j) = c$ for all $(i, j) \in E$, the weight function is said to be constant. Note, that without loss of generality, we can assume $c = 1$.

[1] pointed out that the multiway cut problem is NP-complete even for $|N| = 3, |N_i| = 1$ and constant weight functions. But for the special case of a tree, [2] showed that multiway cut can be solved in polynomial time even for color-dependent weight functions. The corresponding algorithm has time complexity $O(|V|r^2)$. This finally determines the complexity of the partial distribution

problem, associated with the minimization of communication costs in local computation with weight predictable valuation algebras.

C. Semple - M.A. Steel: *Phylogenetics*, Oxford University Press, Oxford UK (2003) 254 pp. Hivatkozsa:[4, 5, 12, 10]

P. 88.: Although Corollary 5.1.8 only applies to two-state characters, Erdős and Székely (1992) developed an extension to arbitrary characters that differs by permitting path to intersect provided certain condition are met.

Suppose we have a phylogenetic X -tree $\mathcal{T} = (T; \phi)$ and a character χ on X . A collection $\mathcal{D}$ of directed paths is an *Erdős-Székely path system* for χ on $\mathcal{T}$ if it satisfies the following two conditions:

- (i) If $P \in \mathcal{D}$ then P connects two leaves $\phi(x)$ and $\phi(y)$ of $\mathcal{T}$ for which $\chi(x) \neq \chi(y)$.
- (ii) Let P and P' be paths in $\mathcal{D}$ that share some edge. Then, P and P' traverse this edge in the same direction and, if $\phi(x)$ and $\phi(y)$ denote the terminal vertices of P and P' , $\chi(x) \neq \chi(y)$.

HIVATKOZÁSOK

- [1] P.L. Erdős - P. Frankl - G.O.H. Katona: Intersecting Sperner families and their convex hulls, *Combinatorica* **4** (1984), 21-34.
- [2] P.L. Erdős - P. Frankl - G.O.H. Katona: Extremal hypergraphs problems and convex hulls, *Combinatorica* **5** (1985), 11-26.
- [3] P.L. Erdős - G.O.H. Katona: Convex hulls of more-part Sperner families, *Graphs and Combin.* **2** (1986), 123-134.
- [4] P.L. Erdős - L.A. Székely: Applications of antilexicographical order I. An enumerative theory of trees, *Advances in Applied Mathematics* **10** (1989), 488-496.
- [5] P.L. Erdős - L. A. Székely: Evolutionary trees: an integer multicommodity max-flow – min-cut theorem, *Advances in Appl. Math* **13** (1992) 375-389.
- [6] P.L. Erdős - L.A. Székely: Algorithms and min-max theorems for certain multiway cuts, *Integer Programming and Combinatorial Optimization* (Proc. of a Conf. held at Carnegie Mellon University, May 25-27, 1992, by the Math. Programming Society, ed. by E. Balas, G. Cornuéjols, R. Kannan) 334-345.
- [7] P.L. Erdős - U. Faigle - W. Kern: A group-theoretic setting for some intersecting Sperner families, *Combinatorics, Probability and Computing* **1** (1992), 323-334.
- [8] M.A. Steel - M.D. Hendy - L.A. Székely - P.L. Erdős: Spectral analysis and a closest tree method for genetic sequences, *Appl. Math. Letters* **5** (1992), 63-67.
- [9] L.A. Székely - M. Steel - P.L. Erdős: Fourier calculus on evolutionary trees, *Advances in Appl. Math* **14** (1993), 200-216.
- [10] P.L. Erdős - L. A. Székely: Counting bichromatic evolutionary trees, *Discrete Applied Mathematics* **47** (1993), 1-8.
- [11] M.A. Steel - L.A. Székely - P.L. Erdős - P. Waddell: A complete family of phylogenetic invariants for any number of taxa, *NZ Journal of Botany*, **31** (1993), 289-296.
- [12] P.L. Erdős - L. A. Székely: On weighted multiway cuts in trees, *Mathematical Programming* **65** (1994), 93-105.
- [13] R. Ahlswede - P.L. Erdős - Niall Graham: A splitting property of maximal antichains, *Combinatorica* **15** (1995), 475-480.
- [14] P.L. Erdős- U. Faigle - W. Kern: On the average rank of LYM-sets, *Discrete Math.* **144** (1995), 11-22.
- [15] P.L. Erdős: Splitting property in infinite posets, *Discrete Math.* **163** (1997), 251–256.
- [16] P.L. Erdős - M.A. Steel - L.A. Székely- T.J. Warnow: Local quartet splits of a binary tree infer all quartet splits via one dyadic inference rule, *Computers and Artificial Intelligence* **16** (1997), 217-227.
- [17] P.L. Erdős - K. Rice - M.A. Steel - L.A. Székely- T.J. Warnow: The Short Quartet Method, to appear in Math. Modelling and Sci. Computing Special Issue of the papers presented at the

Computational Biology sessions at the 11th ICMCM, March 31 - April 2, 1997, Georgetown University Conference Center, Washington, D.C., USA.

- [18] P.L. Erdős - M.A. Steel - L.A. Székely- T.J. Warnow: Constructing big trees from short sequences, *Automata, Languages and Programming* 24th International Colloquium, ICALP'97, Bologna, Italy, July 7 - 11, 1997, (P. Degano,; R. Gorrieri, A. Marchetti-Spaccamela, Eds.) Proceedings (Lecture Notes in Computer Science. Vol. 1256) (1997), 827-837.
- [19] P.L. Erdős - M.A. Steel - L.A. Székely - T.J. Warnow: A few logs suffice to build (almost) all trees (I), *Random Structures and Algorithms* **14** (1999), 153-184.
- [20] P.L. Erdős - M.A. Steel - L.A. Székely- T.J. Warnow: A few logs suffice to build (almost) all trees (II), *Theoretical Computer Science*, **221** (1-2) (1999), 77-118.