produces agreement (see [2] and references therein). The approximability of the MAST and related problem has been studied in [3] and references therein.

Cross-References

- ▶ Maximum Agreement Subtree (of 2 Binary Trees)
- ▶ Maximum Agreement Supertree
- ▶ Maximum Compatible Tree

Acknowledgments This work was supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine.

Recommended Reading

1. Amir A, Keselman D (1997) Maximum agreement subtree in a set of evolutionary trees: metrics and efficient algorithms. SIAM J Comput 26(6):1656–1669
2. Berry V, Nicolas F (2006) Improved parameterized complexity of the maximum agreement subtree and maximum compatible tree problems. IEEE/ACM Trans Comput Biol Bioinform 3(3):289–302
3. Berry V, Guillemot S, Nicolas F, Paul C (2005) On the approximation of computing evolutionary trees. In: COCOON, Kunming, pp 115–125
4. Bryand D (1997) Building trees, hunting for trees, and comparing trees: theory and methods in phylogenetic analysis. Ph.D. thesis, Department of Mathematics, University of Canterbury
5. Cole R, Farach-Colton M, Hariharan R, Przytycka T, Thorup M (2001) An $o(n \log n)$ algorithm for the maximum agreement subtree problem for binary trees. SIAM J Comput 1385–1404
6. Farach M, Przytycka TM, Thorup M (1995) On the agreement of many trees. Inf Process Lett 55(6):297–301
7. Finden CR, Gordon AD (1985) Obtaining common pruned trees. J Classif 2:255–276
8. Kao M-Y, Lam T-W, Sung W-K, Ting H-F (2001) An even faster and more unifying algorithm for comparing trees via unbalanced bipartite matchings. J Algorithms 40(2):212–233
9. Lee C-M, Hung L-J, Chang M-S, Tang C-Y (2004) An improved algorithm for the maximum agreement subtree problem. In: BIBE, Taichung, p 533
10. Przytycka TM (1998) Transforming rooted agreement into unrooted agreement. J Comput Biol 5(2):335–349
11. Steel MA, Warnow T (1993) Kaikoura tree theorems: computing the maximum agreement subtree. Inf Process Lett 48(2):77–82

Maximum Agreement Supertree

Jesper Jansson[1] and Wing-Kin Sung[2]
[1]Laboratory of Mathematical Bioinformatics, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, Japan
[2]Department of Computer Science, National University of Singapore, Singapore, Singapore

Keywords

Fixed-parameter tractability; Maximum agreement supertree; NP-hardness; Phylogenetic tree; Rooted triplet

Years and Authors of Summarized Original Work

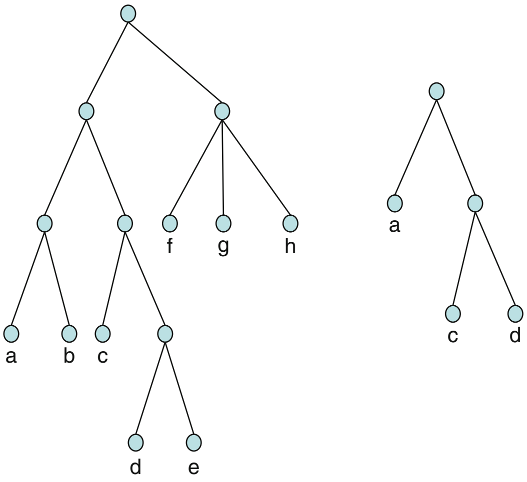
2005; Jansson, Ng, Sadakane, Sung
2007; Berry, Nicolas
2010; Guillemot, Berry
2011; Hoang, Sung

Problem Definition

A *phylogenetic tree* is a rooted, unordered tree whose leaves are distinctly labeled and whose internal nodes have degree at least two. By distinctly labeled, we mean that no two leaves in the tree have the same label. Let T be a phylogenetic tree with a leaf label set S. For any subset S' of S, *the topological restriction of* T *to* S' (denoted by $T \mid S'$) is the tree obtained from T by deleting all nodes which are not on any path from the root to a leaf in S' along with their incident edges and then contracting every edge between a node having just one child and its child. See Fig. 1 for an illustration. For any phylogenetic tree T, denote its set of leaf labels by $\Lambda(T)$.

The *maximum agreement supertree problem (MASP)* [12] is defined as follows.

Problem 1 Let $\mathcal{T} = \{T_1, T_2, \ldots, T_k\}$ be an input set of phylogenetic trees, where the sets $\Lambda(T_i)$ may overlap. The maximum agreement supertree problem (MASP) asks for a phylogenetic



Maximum Agreement Supertree, Fig. 1 Let T be the phylogenetic tree on the *left*. Then $T \mid \{a, c, d\}$ is the phylogenetic tree shown on the *right*

tree Q with leaf label set $\Lambda(Q) \subseteq \bigcup_{T_i \in \mathcal{T}} \Lambda(T_i)$ such that $|\Lambda(Q)|$ is maximized and for each $T_i \in \mathcal{T}$, it holds that $T_i \mid \Lambda(Q)$ is isomorphic to $Q \mid \Lambda(T_i)$.

The following notation is used below: $n = \left|\bigcup_{T_i \in \mathcal{T}} \Lambda(T_i)\right|$, $k = |\mathcal{T}|$, and $D = \max_{T_i \in \mathcal{T}}\{\deg(T_i)\}$, where $\deg(T_i)$ is the degree of T_i (i.e., the maximum number of children of any node belonging to T_i).

A problem related to MASP is the *maximum compatible supertree problem (MCSP)* [2]:

Problem 2 Let $\mathcal{T} = \{T_1, T_2, \ldots, T_k\}$ be an input set of phylogenetic trees, where the sets $\Lambda(T_i)$ may overlap. The maximum compatible supertree problem (MCSP) asks for a phylogenetic tree W with leaf label set $\Lambda(W) \subseteq \bigcup_{T_i \in \mathcal{T}} \Lambda(T_i)$ such that $|\Lambda(W)|$ is maximized and for each $T_i \in \mathcal{T}$, it holds that $T_i \mid \Lambda(W)$ can be obtained from $W \mid \Lambda(T_i)$ by applying a series of edge contractions.

For information about MCSP, refer to [2, 11].

Key Results

The special case of the maximum agreement supertree problem in which $\Lambda(T_1) = \Lambda(T_2) \ldots = \Lambda(T_k)$ has been well studied in the literature and is also known as the *maximum agreement subtree problem (MAST)*. By utilizing known results for MAST, several results can be obtained for various special cases of MASP. Firstly, it is known that MAST can be solved in $O(\sqrt{D}n\log(2n/D))$ time when $k = 2$ (see [13]) or in $O(kn^3 + n^D)$ time when $k \geq 3$ (see [4, 6]), which leads to the following theorems.

Theorem 1 ([12]) *When $k = 2$, MASP can be solved in $O(T_{MAST} + n)$ time, where T_{MAST} is the time required to solve MAST for two $O(n)$-leaf trees. Note that $T_{MAST} = O(\sqrt{D}\, n \log(2n/D))$.*

Theorem 2 ([2]) *For any fixed $k \geq 3$, if every leaf appears in either 1 or k trees, MASP can be solved in $O(T'_{MAST} + kn)$ time, where T'_{MAST} is the time required to solve MAST for $\{T_1|L, T_2|L, \ldots, T_k|L\}$, where $L = \bigcap_{T_i \in \mathcal{T}} \Lambda(T_i)$. Note that $T'_{MAST} = O(k|L|^3 + |L|^D)$.*

On the negative side, the maximum agreement supertree problem is NP-hard in general, as shown by the next theorem. (A *rooted triplet* is a binary phylogenetic tree with exactly three leaves.)

Theorem 3 ([2, 12]) *For any fixed $k \geq 3$, MASP with unbounded D is NP-hard. Furthermore, MASP with unbounded k remains NP-hard even if restricted to rooted triplets, i.e., $D = 2$.*

The inapproximability results for MAST by Hein et al. [9] and Gąsieniec et al. [7] immediately carry over to MASP with unbounded D as follows.

Theorem 4 ([2, 12]) *cannot be approximated within a factor of $2^{\log^\delta n}$ in polynomial time for any constant $\delta < 1$, unless NP $\subseteq$ DTIME[$2^{\text{polylog}\, n}$], even when restricted to $k = 3$. Also, MASP cannot be approximated within a factor of n^ε for any constant ε where $0 \leq \varepsilon < \frac{1}{9}$ in polynomial time unless P = NP, even for instances containing only trees of height 2.*

Although MASP is difficult to approximate in polynomial time, a simple approximation algorithm based on a technique from [1] achieves an approximation factor that is close to the bounds given in Theorem 4.

Theorem 5 **([12])** *MASP can be approximated within a factor of* $\frac{n}{\log n}$ *in* $O(n^2) \cdot \min\{O(k \cdot (\log\log n)^2), O(k + \log n \cdot \log\log n)\}$ *time. MASP restricted to rooted triplets can be approximated within a factor of* $\frac{n}{\log n}$ *in* $O(k + n^2 \log^2 n)$ *time.*

Fixed-parameter tractable algorithms for solving MASP also exist. In particular, for *binary* phylogenetic trees, Jansson et al. [12] first gave an $O(k(2n^2)^{3k^2})$-time algorithm. Later, Guillemot and Berry [8] improved the time complexity to $O((8n)^k)$. Hoang and Sung [11] further improved the time complexity to $O((6n)^k)$, as summarized in Theorem 6.

Theorem 6 **([11])** *MASP restricted to* $D = 2$ *can be solved in* $O((6n)^k)$ *time.*

For the case where each tree in $\mathcal{T}$ has degree at most D, Hoang and Sung [11] gave the following fixed-parameter polynomial-time solution.

Theorem 7 **([11])** *MASP restricted to phylogenetic trees of degree at most* D *can be solved in* $O((kD)^{kD+3}(2n)^k)$ *time.*

For unbounded n, k, and D, Guillemot and Berry [8] proposed a solution that is efficient when the input trees are similar.

Theorem 8 **([8])** *MASP can be solved in* $O((2k)^p k n^2)$ *time, where* p *is an upper bound on the number of leaves that are missing from* $\bigcup_{T_i \in \mathcal{T}} \Lambda(T_i)$ *in a MASP solution.*

Applications

One challenge in phylogenetics is to develop good methods for merging a collection of phylogenetic trees on overlapping sets of taxa into a single supertree so that no (or as little as possible) branching information is lost. Ideally, the resulting supertree can then be used to deduce evolutionary relationships between taxa which do not occur together in any one of the input trees. Supertree methods are useful because most individual studies investigate relatively few taxa [15] and because sample bias leads to certain taxa being studied much more frequently than others [3]. Also, supertree methods can combine trees constructed for different types of data or under different models of evolution. Furthermore, although computationally expensive methods for constructing reliable phylogenetic trees are infeasible for large sets of taxa, they can be applied to obtain highly accurate trees for smaller, overlapping subsets of the taxa which may then be merged using computationally less intense, supertree-based techniques (see, e.g., [5, 10, 14]).

Since the set of trees which is to be combined may in practice contain contradictory branching structure (e.g., if the trees have been constructed from data originating from different genes or if the experimental data contains errors), a supertree method needs to specify how to resolve conflicts. One intuitive idea is to identify and remove a smallest possible subset of the taxa so that the remaining taxa can be combined without conflicts. In this way, one would get an indication of which ancestral relationships can be regarded as resolved and which taxa need to be subjected to further experiments. The above biological problem can be formalized as MASP.

Open Problems

An open problem is to improve the time complexity of the currently fastest algorithms for solving MASP. Moreover, the existing fixed-parameter polynomial-time algorithms for MASP are not practical, so it could be useful to provide heuristics that work well on real data.

Cross-References

- ▶ Maximum Agreement Subtree (of 2 Binary Trees)
- ▶ Maximum Agreement Subtree (of 3 or More Trees)
- ▶ Maximum Compatible Tree

Acknowledgments JJ was funded by the Hakubi Project at Kyoto University and KAKENHI grant number 26330014.

Recommended Reading

1. Akutsu T, Halldórsson MM (2000) On the approximation of largest common subtrees and largest common point sets. Theor Comput Sci 233(1–2): 33–50
2. Berry V, Nicolas F (2007) Maximum agreement and compatible supertrees. J Discret Algorithms 5(3):564–591
3. Bininda-Emonds ORP, Gittleman JL, Steel MA (2002) The (super)tree of life: procedures, problems, and prospects. Annu Rev Ecol Syst 33: 265–289
4. Bryant D (1997) Building trees, hunting for trees, and comparing trees: theory and methods in phylogenetic analysis. PhD thesis, University of Canterbury, Christchurch
5. Chor B, Hendy M, Penny D (2007) Analytic solutions for three taxon ML trees with variable rates across sites. Discret Appl Math 155(6–7): 750–758
6. Farach M, Przytycka T, Thorup M (1995) On the agreement of many trees. Inf Process Lett 55(6):297–301
7. Gąsieniec L, Jansson J, Lingas A, Östlin A (1999) On the complexity of constructing evolutionary trees. J Comb Optim 3(2–3):183–197
8. Guillemot S, Berry V (2010) Fixed-parameter tractability of the maximum agreement supertree problem. IEEE/ACM Trans Comput Biol Bioinform 7(2):342–353
9. Hein J, Jiang T, Wang L, Zhang K (1996) On the complexity of comparing evolutionary trees. Discret Appl Math 71(1–3):153–169
10. Henzinger MR, King V, Warnow T (1999) Constructing a tree from homeomorphic subtrees, with applications to computational evolutionary biology. Algorithmica 24(1):1–13
11. Hoang VT, Sung W-K (2011) Improved algorithms for maximum agreement and compatible supertrees. Algorithmica 59(2):195–214
12. Jansson J, Ng JHK, Sadakane K, Sung W-K (2005) Rooted maximum agreement supertrees. Algorithmica 43(4):293–307
13. Kao M-Y, Lam T-W, Sung W-K, Ting H-F (2001) An even faster and more unifying algorithm for comparing trees via unbalanced bipartite matchings. J Algorithms 40(2):212–233
14. Kearney P (2002) Phylogenetics and the quartet method. In: Current topics in computational molecular biology. MIT, Cambridge pp 111–133
15. Sanderson MJ, Purvis A, Henze C (1998) Phylogenetic supertrees: assembling the trees of life. TRENDS Ecol Evol 13(3):105–109

Maximum Cardinality Stable Matchings

Eric McDermid
Cedar Park, TX, USA

Keywords

Approximation algorithm; Lower bounds; Matching; NP-hard; Preferences; Stability; Ties; UGC-hard; Upper bounds

Years and Authors of Summarized Original Work

2002; Manlove, Irving, Iwama, Miyazaki, Morita
2003; Halld'orsson, Iwama, Miyazaki, Yanagisawa
2004; Halld'orsson, Iwama, Miyazaki, Yanagisawa
2004; Iwama, Miyazaki, Okamoto
2007; Halld'orsson, Iwama, Miyazaki, Yanagisawa
2007; Iwama, Miyazaki, Yamauchi
2007; Yanagisawa
2008; Irving, Manlove
2008; Iwama, Miyazaki, Yamauchi
2009; McDermid
2011; Kir'aly
2013; Kir'aly
2014; Huang and Kavitha
2014; Iwama, Miyazaki, and Yanagisawa
2014; Paluch
2014; Radnai
2015; Dean, Jalasutram

Problem Definition

The input to an instance of the classical *stable marriage problem* consists of a set of n men and n women. Additionally, each person provides a strictly ordered preference list of the opposite set. The goal is to find a complete matching of men to women that is also *stable*, i.e., a matching having the property that there does not exist a