

Trees in Many Contexts

Heather Smith Blake, Miklós Bóna, Éva Czabarka,
Stephan Wagner, and Hua Wang

The authors of this piece are organizers of the AMS 2022 Mathematics Research Communities summer conference *Trees in Many Contexts*, one of four topical research conferences offered this year that are focused on collaborative research and professional development for early-career mathematicians. Additional information can be found at <https://www.ams.org/programs/research-communities/2022MRC-Trees>. Applications are open until February 15, 2022.

Trees are one of the most fundamental classes of graphs. They play a key role in graph theory and combinatorics, but also occur in many other contexts within mathematics as well as other sciences, such as biology, chemistry, and computer science.

Trees have a very simple structure, so many graph-theoretical problems that are otherwise very difficult can be solved for trees, and algorithmic questions that are computationally hard for general graphs often have simple efficient solutions. At the same time, there are also notoriously hard problems that specifically concern trees, for example *Ringel's conjecture*, which states that the complete graph K_{2n-1} can be decomposed into $2n - 1$ edge-disjoint

copies of any given tree T with n vertices. This conjecture was only verified very recently for large n [9].

An example of a problem on trees that is completely open is *Graham's tree reconstruction conjecture*: starting from a tree T , take the line graph $L(T)$ (the graph whose vertices are the edges of T and where two vertices are adjacent if and only if the corresponding edges are adjacent in T) and its iterates $L(T), L(L(T)), L(L(L(T))), \dots$. Graham's tree reconstruction conjecture states that the sequence of orders $|T|, |L(T)|, |L(L(T))|, \dots$ uniquely determines T . Not much is known about this conjecture. Recent progress [3] shows that the number of trees with n vertices that are uniquely determined is at least $e^{\Omega((\log n)^{3/2})}$, but this is still far from the number of different trees, which grows exponentially with n .

These examples already illustrate the variety of interesting problems that deal with trees, and also show that trees are only superficially simple and easy to understand.

Trees in Biology

In biology, trees occur prominently as *phylogenetic trees* that illustrate the evolutionary relationships among different species. The species correspond to the leaves of the tree, and the root (if there is one) to the most recent common ancestor. Mathematical problems involving phylogenetic trees are often motivated by computational applications.

A typical example to illustrate the type of questions that arise in this context concerns *maximum agreement subtrees*. Here, one considers leaf-labeled binary trees, where the leaves are labeled $1, 2, \dots, n$ (one can think of them as representing n different species). A binary tree can be restricted to a subset of leaves S in a natural way: one takes the smallest subtree that contains all these leaves, then one suppresses all vertices of degree 2 to obtain a new binary tree.

Two leaf-labeled trees T_1 and T_2 with n vertices are said to *agree* on a subset S of $\{1, 2, \dots, n\}$ if the respective

Heather Smith Blake is an assistant professor of mathematics at Davidson College. Her email address is hsblake@davidson.edu.

Miklós Bóna is a professor of mathematics at the University of Florida. His email address is bona@ufl.edu.

Éva Czabarka is a professor of mathematics at the University of South Carolina. Her email address is czabarka@math.sc.edu.

Stephan Wagner is a professor of mathematics at Uppsala University. His email address is stephan.wagner@math.uu.se.

Hua Wang is a professor of mathematics at Georgia Southern University. His email address is hwang@georgiasouthern.edu.

For permission to reprint this article, please contact:
reprint-permission@ams.org.

DOI: <https://doi.org/10.1090/noti2411>

restrictions to S are the same (i.e., there is a label-preserving isomorphism). For example, the two trees in Figure 1 agree on the set $\{1, 3, 5, 6\}$. The maximum size of a set S on which T_1 and T_2 agree is denoted $\text{MAST}(T_1, T_2)$. This quantity is similar, for example, to the longest common subsequence of two strings or permutations. It was an open question until recently how small $\text{MAST}(T_1, T_2)$ could be in terms of the number of leaves n . In his recent paper [6], Markin established that the minimum is in fact of order $\Theta(\log n)$ by improving the previous lower bound that was of order $\sqrt{\log n}$ [7].

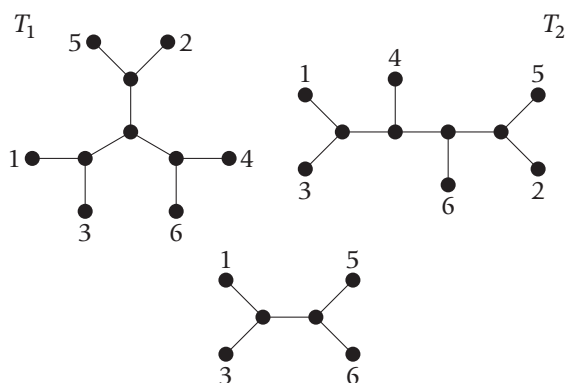


Figure 1. Two trees T_1 and T_2 and their restriction to $\{1, 3, 5, 6\}$.

For pairs of *random trees*, the order of magnitude of the average size of a maximum agreement subtree is still unknown. Upper bounds of $O(\sqrt{n})$ hold for both standard models in phylogenetics, the uniform model and the Yule-Harding model. On the other hand, Bernstein, Ho, Long, Steel, St. John, and Sullivant [1] provided lower bounds of $\Omega(n^{1/8})$ (uniform model) and $\Omega(n^{0.344})$ (Yule-Harding). Simulations suggest that the true exponent is close to $\frac{1}{2}$; see [2]. For random trees of the same shape (but different labels), it was shown recently [8] that the order of the expected size is $\Theta(\sqrt{n})$.

Trees in Chemistry

Chemistry has influenced combinatorial mathematics for a long time, an early example being the enumeration of chemical compounds, which was an explicit motivation in Pólya's ground-breaking work [10] on what is now known as *Pólya's enumeration method*. Trees play a prominent role in this context as models for acyclic molecules. For example, there are three skeletal isomers of pentane C_5H_{12} , which correspond to the isomorphism classes of trees with five vertices, as shown in Figure 2.

The field of *chemical graph theory* has seen a lot of growth in recent years, which is also evidenced by the inclusion of two new MSC codes in the 2020 edition: *chemical graph theory* 05C92 and *graphical indices* 05C09. Graphical indices

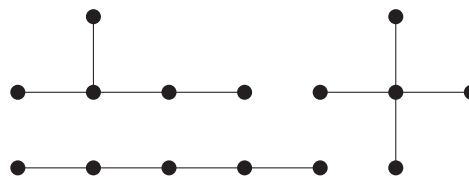


Figure 2. The three trees of order 5.

(also called *topological indices* after an early example that is now known as Hosoya index) are invariants (usually integers or real numbers) associated with graphs. In many instances, it has been observed that these indices are well correlated with physico-chemical properties of the corresponding molecules.

A prominent instance is the *Wiener index* of a graph, which is defined as the sum of the distances between all pairs of vertices. It was introduced by Harry Wiener in his 1947 paper [13] as the *path number* in connection with paraffin boiling points. As Wiener notices in passing, it can be calculated for trees as follows: for each edge e , take the product of the cardinalities of the two components that result when e is removed. Then sum all the resulting products. It is easy to see that the product represents the number of times e occurs on a (shortest) path between two vertices, so this sum is indeed equal to the Wiener index. This simple—yet elegant and useful—property of the Wiener index is an early example of a theorem in chemical graph theory that actually appeared in a chemistry journal.

There are many different aspects to current chemical graph theory, but most commonly, researchers are interested in bounds for various indices associated with graphs, relations between different indices, and extremal questions where the maximum or minimum of an index for a specific family of graphs (for instance, all trees with n vertices) are sought. Trees traditionally play a prominent role in this context, as do various types of tree-like graphs that are relevant in chemistry, e.g., graphs with a limited number of cycles or cycles of specific lengths such as hexagons. Certain structures occur repeatedly in this context; unsurprisingly, for instance, path and star are typically the trees for which minimum and maximum of a graphical index are attained. Another example is *greedy trees* that frequently occur when degree restrictions are imposed; see for example [12, Sections 2.6.2 and 3.2.1].

Trees in Computer Science

Trees are useful as a data structure and form the basis of many algorithms; see for example [5, Section 2.3]. For example, trees store hierarchical data and can be used for searching and sorting. *Binary search trees* are a typical example of a tree data structure, and they are also connected to the well-known *Quicksort* algorithm. In a binary search tree, labels in the left root branch are all smaller than

the root label while labels in the right root branch are all greater. The same applies to the two branches at all other vertices.

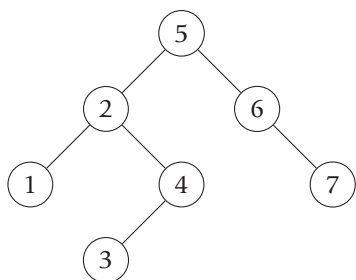


Figure 3. A binary search tree.

Trees have been studied for a long time in the context of worst-case and average-case analysis of algorithms; see for example [11, Chapter 6]. Parameters like the *height* (greatest distance from the root to a leaf) are measures for the performance of algorithms. Within this context, enumerative problems concerning trees and tree-like structures as well as properties of *random trees* are often of interest, especially when average-case performance of algorithms is considered. There are many different ways to generate trees at random, and the structural properties often differ depending on the specific model of randomness that is used. The simplest and most straightforward way to generate trees at random is perhaps *uniform models*, in which a tree is selected uniformly at random from all trees with n vertices from a given family.

A classical way to generate trees by a random process is *Galton-Watson trees*: these trees start with a root and a given offspring distribution. In the k th step, all vertices of generation k independently produce offspring according to that distribution. Galton-Watson trees and uniformly random trees are models where typical trees with n vertices have height of order \sqrt{n} .

Other models produce trees where the height is typically only of logarithmic order: examples include random binary search trees (whose probabilistic model is in fact equivalent to the aforementioned Yule-Harding model in phylogenetics) and *recursive trees*, where one starts with a single vertex and the n th vertex is attached to one of the previous $n - 1$ vertices uniformly at random. The height is only one of the many structural properties of trees that have been studied for many different types of random trees. Very precise results not only on the order of magnitude, but also the distribution, are often available [4].

Conclusion

This brief introduction aims to show in what different contexts trees and tree structures occur and what mathematical questions around trees are of interest in different fields.

Its intention is also to give a flavor of the type of problems that will be investigated at the upcoming AMS MRC *Trees in Many Contexts*.

ACKNOWLEDGMENT. Stephan Wagner is supported by the Knut and Alice Wallenberg Foundation.

References

- [1] Daniel Irving Bernstein, Lam Si Tung Ho, Colby Long, Mike Steel, Katherine St. John, and Seth Sullivant, *Bounds on the expected size of the maximum agreement subtree*, SIAM J. Discrete Math. **29** (2015), no. 4, 2065–2074, DOI 10.1137/140997750. MR3416130
- [2] David Bryant, Andy McKenzie, and Mike Steel, *The size of a maximum agreement subtree for random binary trees*, Bioconsensus (Piscataway, NJ, 2000/2001), DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 61, Amer. Math. Soc., Providence, RI, 2003, pp. 55–65, DOI 10.1090/dimacs/061/04. MR1982419
- [3] Joshua Cooper, Bill Kay, and Anton Swifton, *Graham's tree reconstruction conjecture and a Waring-type problem on partitions*, J. Comb. **9** (2018), no. 3, 469–488, DOI 10.4310/JOC.2018.v9.n3.a3. MR3809644
- [4] Michael Drmota, *Random trees: An interplay between combinatorics and probability*, SpringerWienNewYork, Vienna, 2009, DOI 10.1007/978-3-211-75357-6. MR2484382
- [5] Donald E. Knuth, *The art of computer programming. Vol. 1: Fundamental algorithms*, Addison-Wesley, Reading, MA, 1997. Third edition [of MR0286317]. MR3077152
- [6] Alexey Markin, *On the extremal maximum agreement subtree problem*, Discrete Appl. Math. **285** (2020), 612–620, DOI 10.1016/j.dam.2020.07.007. MR4124794
- [7] Daniel M. Martin and Bhalchandra D. Thatte, *The maximum agreement subtree problem*, Discrete Appl. Math. **161** (2013), no. 13–14, 1805–1817, DOI 10.1016/j.dam.2013.02.037. MR3056987
- [8] Pratik Misra and Seth Sullivant, *Bounds on the expected size of the maximum agreement subtree for a given tree shape*, SIAM J. Discrete Math. **33** (2019), no. 4, 2316–2325, DOI 10.1137/18M1213695. MR4036089
- [9] R. Montgomery, A. Pokrovskiy, and B. Sudakov, *A proof of Ringel's conjecture*, Geom. Funct. Anal. **31** (2021), no. 3, 663–720, DOI 10.1007/s00039-021-00576-2. MR4311581
- [10] G. Pólya, *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen* (German), Acta Math. **68** (1937), no. 1, 145–254, DOI 10.1007/BF02546665. MR1577579
- [11] Robert Sedgewick and Philippe Flajolet, *An introduction to the analysis of algorithms*, Addison-Wesley-Longman, 1996.
- [12] Stephan Wagner and Hua Wang, *Introduction to chemical graph theory*, Discrete Mathematics and its Applications (Boca Raton), CRC Press, Boca Raton, FL, 2019. MR3837106
- [13] H. Wiener, *Structural determination of paraffin boiling points*, J. Am. Chem. Soc. **69** (1947), 17–20.