Good Will Hunting's Problem: Counting Homeomorphically Irreducible Trees

Ira M. Gessel

Department of Mathematics
Brandeis University

Brandeis University Combinatorics Seminar
September 18, 2018

Good Will Hunting's Problem

In the movie *Good Will Hunting*, Professor Lambeau says:

Good Will Hunting's Problem

In the movie *Good Will Hunting*, Professor Lambeau says:

My colleagues and I have conferred, and there is a problem on the board right now that took us more than two years to prove. Let this be said: the gauntlet has been thrown down. But the faculty have answered, and answered with vigor.

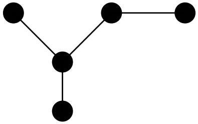
Good Will Hunting's Problem

In the movie *Good Will Hunting*, Professor Lambeau says:

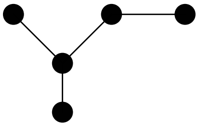
My colleagues and I have conferred, and there is a problem on the board right now that took us more than two years to prove. Let this be said: the gauntlet has been thrown down. But the faculty have answered, and answered with vigor.

The problem seems to be "Draw all the homeomorphically irreducible trees with $n = 10$."

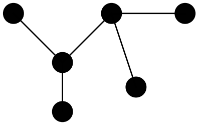
A tree is a connected graph without cycles.



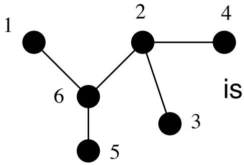
A tree is a connected graph without cycles.



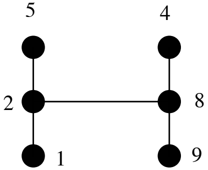
A tree is homeomorphically irreducible if it has no vertices of degree 2.



We want to count unlabeled trees, which more formally are isomorphism classes of trees. For example,

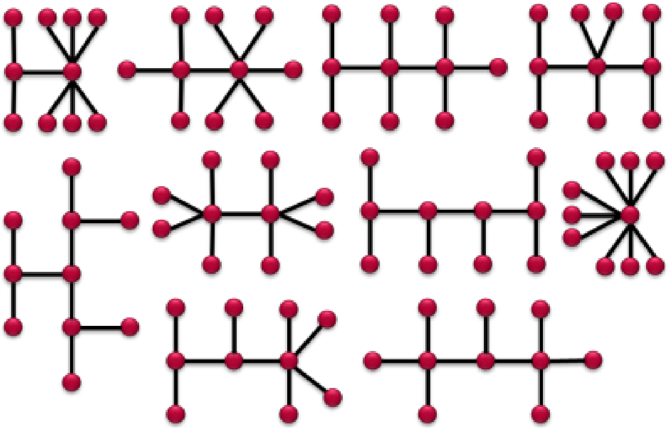


is isomorphic to



The answer to Good Will Hunting’s problem is that there are 10 unlabeled homeomorphically irreducible trees with 10 vertices.

The answer to Good Will Hunting's problem is that there are 10 unlabeled homeomorphically irreducible trees with 10 vertices.



In the movie Will just draws them (actually he's interrupted after drawing 8 of them). It's not hard to show that these are all different.

In the movie Will just draws them (actually he's interrupted after drawing 8 of them). It's not hard to show that these are all different.

How do we know we haven't missed any?

In the movie Will just draws them (actually he's interrupted after drawing 8 of them). It's not hard to show that these are all different.

How do we know we haven't missed any?

Let n_i be the number of vertices of degree i. Then $\sum_i n_i = 10$.

In the movie Will just draws them (actually he's interrupted after drawing 8 of them). It's not hard to show that these are all different.

How do we know we haven't missed any?

Let n_i be the number of vertices of degree i. Then $\sum_i n_i = 10$.

Also, in any graph the sum of the degrees is twice the number of edges. A tree with m vertices has $m - 1$ edges. So $\sum_i i n_i = 18$.

In the movie Will just draws them (actually he's interrupted after drawing 8 of them). It's not hard to show that these are all different.

How do we know we haven't missed any?

Let n_i be the number of vertices of degree i. Then $\sum_i n_i = 10$.

Also, in any graph the sum of the degrees is twice the number of edges. A tree with m vertices has $m - 1$ edges. So $\sum_i i n_i = 18$.

It's convenient to eliminate n_1 from these equations. Subtracting the first from the second gives

$$8n_9 + 7n_8 + 6n_7 + 5n_6 + 4n_5 + 3n_4 + 2n_3 + n_2 = 8.$$

Then n_9, n_8, n_7, and n_6 must all be 0 or 1, and at most one of them can be 1, and n_2 must be 0. . . . It's not hard to check all the possibilities.

A more interesting problem

A much more interesting problem is to find the number of homeomorphically irreducible trees with *n* vertices for all *n*.

A more interesting problem

A much more interesting problem is to find the number of homeomorphically irreducible trees with *n* vertices for all *n*.

But what does this mean?

A more interesting problem

A much more interesting problem is to find the number of homeomorphically irreducible trees with n vertices for all n.

But what does this mean?

Ideally, we'd like a nice formula for the number of homeomorphically irreducible trees with n vertices, like Cayley's formula n^{n-2} for the number of labeled trees with vertex set $\{1, 2, \ldots, n\}$.

A more interesting problem

A much more interesting problem is to find the number of homeomorphically irreducible trees with n vertices for all n.

But what does this mean?

Ideally, we'd like a nice formula for the number of homeomorphically irreducible trees with n vertices, like Cayley's formula n^{n-2} for the number of labeled trees with vertex set $\{1, 2, \ldots, n\}$.

But there is no such formula.

A more interesting problem

A much more interesting problem is to find the number of homeomorphically irreducible trees with n vertices for all n.

But what does this mean?

Ideally, we'd like a nice formula for the number of homeomorphically irreducible trees with n vertices, like Cayley's formula n^{n-2} for the number of labeled trees with vertex set $\{1, 2, \ldots, n\}$.

But there is no such formula.

Alternatively, we might hope for an efficient algorithm for computing the number t_n of homeomorphically irreducible trees with n vertices.

A more interesting problem

A much more interesting problem is to find the number of homeomorphically irreducible trees with n vertices for all n.

But what does this mean?

Ideally, we'd like a nice formula for the number of homeomorphically irreducible trees with n vertices, like Cayley's formula n^{n-2} for the number of labeled trees with vertex set $\{1, 2, \ldots, n\}$.

But there is no such formula.

Alternatively, we might hope for an efficient algorithm for computing the number t_n of homeomorphically irreducible trees with n vertices.

We can do better than this.

We will find a (reasonably) simple equation for the generating function

$$\sum_{n=1}^{\infty} t_n x^n$$

from which the coefficients can be computed efficiently.

We will find a (reasonably) simple equation for the generating function

$$\sum_{n=1}^{\infty} t_n x^n$$

from which the coefficients can be computed efficiently.

First we'll do this for trees without the requirement of homeomorphic irreducibility; then counting homeomorphically irreducible trees is a straightforward variation.

We will find a (reasonably) simple equation for the generating function

$$\sum_{n=1}^{\infty} t_n x^n$$

from which the coefficients can be computed efficiently.

First we'll do this for trees without the requirement of homeomorphic irreducibility; then counting homeomorphically irreducible trees is a straightforward variation.

The enumeration of homeomorphically irreducible trees was accomplished by Frank Harary and Geert Prins, *The number of homeomorphically irreducible trees, and other species*, Acta Math. 101 (1959), 141–162.

We will follow the general outline of their approach, with some simplifications and modernizations.

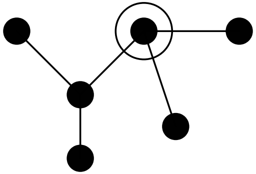
Counting unlabeled trees

We count (unlabeled) trees in two steps.

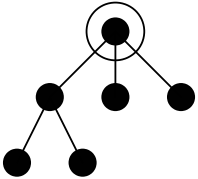
1. First we count rooted trees.
2. Then we reduce the problem of counting unrooted trees to that of counting rooted trees.

Unlabeled trees were first counted by Cayley in 1875; our approach is similar to that of Richard Otter, *The number of trees*, Ann. of Math. (2) 49 (1948), 583–599.

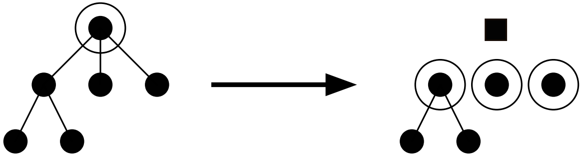
A rooted tree (or more precisely, a vertex-rooted tree) is a tree in which one vertex has been marked as a root:



We often draw rooted trees with the root at the top:



Rooted trees have a recursive decomposition: every (unlabeled) rooted tree may be decomposed into a root and a multiset of rooted trees.



Counting multisets

Let's take a slight detour to count multisets in general. Suppose we have a set $\mathscr{A}$, not necessarily finite. Each element $a \in \mathscr{A}$ has a size $s(a) \in \mathbb{P}$. We give a the weight $x^{s(a)}$.

Counting multisets

Let's take a slight detour to count multisets in general. Suppose we have a set $\mathscr{A}$, not necessarily finite. Each element $a \in \mathscr{A}$ has a size $s(a) \in \mathbb{P}$. We give a the weight $x^{s(a)}$.

We assume that for each integer n, only finitely many elements of $\mathscr{A}$ have size n. Then we define the generating function $A(x)$ for $\mathscr{A}$ to be the formal power series

$$A(x) = \sum_{a \in \mathscr{A}} x^{s(a)}.$$

We define the size of a multiset $\{a_1, a_2, \ldots, a_k\}$ of elements of $\mathscr{A}$ to be $s(a_1) + s(a_2) + \cdots + s(a_k)$. Then the weight of $\{a_1, a_2, \ldots, a_k\}$ is $x^{s(a_1)+s(a_2)+\cdots+s(a_k)}$, the product of the weights of its elements. We would like to find a formula for the generating function (sum of weights) for multisets of elements of $\mathscr{A}$ in terms of $A(x)$.

We define the size of a multiset $\{a_1, a_2, \ldots, a_k\}$ of elements of $\mathscr{A}$ to be $s(a_1) + s(a_2) + \cdots + s(a_k)$. Then the weight of $\{a_1, a_2, \ldots, a_k\}$ is $x^{s(a_1)+s(a_2)+\cdots+s(a_k)}$, the product of the weights of its elements. We would like to find a formula for the generating function (sum of weights) for multisets of elements of $\mathscr{A}$ in terms of $A(x)$.

For multisets made up of copies of a single element $a \in \mathscr{A}$, the generating function is

$$1 + x^{s(a)} + x^{2s(a)} + x^{3s(a)} + \cdots = \frac{1}{1 - x^{s(a)}}$$

We define the size of a multiset $\{a_1, a_2, \ldots, a_k\}$ of elements of $\mathscr{A}$ to be $s(a_1) + s(a_2) + \cdots + s(a_k)$. Then the weight of $\{a_1, a_2, \ldots, a_k\}$ is $x^{s(a_1)+s(a_2)+\cdots+s(a_k)}$, the product of the weights of its elements. We would like to find a formula for the generating function (sum of weights) for multisets of elements of $\mathscr{A}$ in terms of $A(x)$.

For multisets made up of copies of a single element $a \in \mathscr{A}$, the generating function is

$$1 + x^{s(a)} + x^{2s(a)} + x^{3s(a)} + \cdots = \frac{1}{1 - x^{s(a)}}$$

So the generating function for all multisets of elements of $\mathscr{A}$ is

$$\prod_{a \in \mathscr{A}} \frac{1}{1 - x^{s(a)}}$$

It's convenient to write this in another form. The logarithm of this generating function is

$$\begin{aligned}
\log \prod_{a \in \mathscr{A}} \frac{1}{1 - x^{s(a)}} &= \sum_{a \in \mathscr{A}} \log \frac{1}{1 - x^{s(a)}} \\
&= \sum_{a \in \mathscr{A}} \sum_{k=1}^{\infty} \frac{x^{ks(a)}}{k} \\
&= \sum_{k=1}^{\infty} \frac{1}{k} \sum_{a \in \mathscr{A}} (x^k)^{s(a)} \\
&= \sum_{k=1}^{\infty} \frac{1}{k} A(x^k)
\end{aligned}$$

since $A(x) = \sum_{a \in \mathscr{A}} x^{s(a)}$.

Thus if $A(x)$ is the generating function for a set $\mathscr{A}$ then the generating function for multisets of elements of $\mathscr{A}$ is

$$\exp\left(\sum_{k=1}^{\infty} \frac{A(x^k)}{k}\right).$$

Thus if $A(x)$ is the generating function for a set $\mathscr{A}$ then the generating function for multisets of elements of $\mathscr{A}$ is

$$\exp\left(\sum_{k=1}^{\infty} \frac{A(x^k)}{k}\right).$$

Let's write $h_n[A(x)]$ for the generating function for multisets of n elements of $\mathscr{A}$ ("n-multisets") and we write $h[A(x)]$ for $\sum_{n=0}^{\infty} h_n[A(x)]$. Then we have shown

$$h[A(x)] = \exp\left(\sum_{k=1}^{\infty} \frac{A(x^k)}{k}\right).$$

2-multisets

We will also need the generating function for 2-element multisets of elements of $\mathscr{A}$, which is

$$h_2[A(x)] = \frac{1}{2}\left(A(x)^2 + A(x^2)\right).$$

2-multisets

We will also need the generating function for 2-element multisets of elements of $\mathscr{A}$, which is

$$h_2[A(x)] = \frac{1}{2}\left(A(x)^2 + A(x^2)\right).$$

Proof. First, $A(x)^2$ counts ordered pairs of elements of $\mathscr{A}$. So it counts each unordered pair of distinct elements twice and each unordered pair of the same element once.

2-multisets

We will also need the generating function for 2-element multisets of elements of $\mathscr{A}$, which is

$$h_2[A(x)] = \frac{1}{2}\left(A(x)^2 + A(x^2)\right).$$

Proof. First, $A(x)^2$ counts ordered pairs of elements of $\mathscr{A}$. So it counts each unordered pair of distinct elements twice and each unordered pair of the same element once.

Next, $A(x^2)$ counts ordered pairs of the same element of $\mathscr{A}$, so it counts unordered pairs of the same element.

2-multisets

We will also need the generating function for 2-element multisets of elements of $\mathscr{A}$, which is

$$h_2[A(x)] = \frac{1}{2}\left(A(x)^2 + A(x^2)\right).$$

Proof. First, $A(x)^2$ counts ordered pairs of elements of $\mathscr{A}$. So it counts each unordered pair of distinct elements twice and each unordered pair of the same element once.

Next, $A(x^2)$ counts ordered pairs of the same element of $\mathscr{A}$, so it counts unordered pairs of the same element.

Therefore $A(x)^2 + A(x^2)$ counts every unordered pair twice.

Rooted trees

Our decomposition of a rooted tree into a root together with a multiset of rooted trees gives the functional equation for the generating function $R(x)$ for rooted trees, where the size of a rooted tree is the number of vertices:

$$R(x) = xh[R(x)] = x \exp\left(\sum_{k=1}^{\infty} \frac{R(x^k)}{k}\right).$$

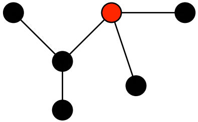
From this equation we can easily compute

$$\begin{aligned} R(x) = {} & x + x^2 + 2x^3 + 4x^4 + 9x^5 + 20x^6 + 48x^7 + 115x^8 + 286x^9 \\ & + 719x^{10} + 1842x^{11} + 4766x^{12} + 12486x^{13} + 32973x^{14} \\ & + 87811x^{15} + 235381x^{16} + 634847x^{17} + 1721159x^{18} \\ & + 4688676x^{19} + 12826228x^{20} + 35221832x^{21} + \cdots \end{aligned}$$

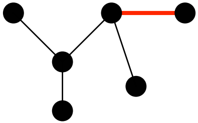
Unrooted trees

To count unrooted trees, we find a formula that expresses the generating function for unrooted trees in terms of the generating function for rooted trees. To do this, we use the dissymmetry theorem of Pierre Leroux. (Otter used a related, but somewhat different, result called the “dissimilarity characteristic theorem”.)

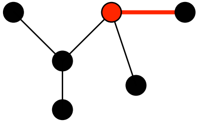
Leroux's theorem relates unrooted trees to trees rooted at a vertex:



trees rooted at an edge:



and trees rooted at a vertex and incident edge:

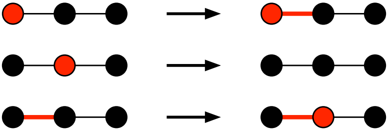


The dissymmetry theorem. Let T be a (labeled) tree. Let $T^{\bullet}$ be the set of rootings of T at a vertex, let T^{-} be the set of rootings of T at an edge, and let $T^{\bullet -}$ be the set of rootings of T at a vertex and incident edge. Then there is a bijection ϕ from $T^{\bullet} \cup T^{-}$ to $\{T\} \cup T^{\bullet -}$.

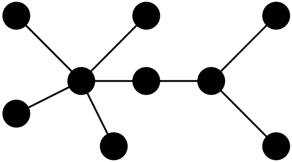
Moreover, ϕ is compatible with isomorphisms of trees; i.e., if $\alpha := T_1 \to T_2$ is an isomorphism of trees than $\phi \circ \alpha = \alpha \circ \phi$.

Thus ϕ gives a corresponding bijection for unlabeled trees.

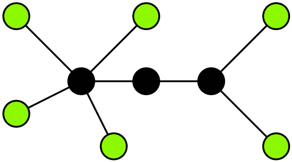
Example:



To prove the dissymmetry theorem, we start with the fact that every tree has either a center edge or a center vertex which is fixed by every automorphism of the tree. The center is obtained by removing every leaf and its incident edge successively until only a vertex or edge remains.



To prove the dissymmetry theorem, we start with the fact that every tree has either a center edge or a center vertex which is fixed by every automorphism of the tree. The center is obtained by removing every leaf and its incident edge successively until only a vertex or edge remains.

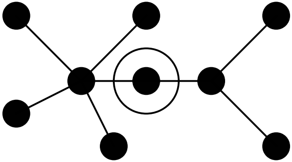


To prove the dissymmetry theorem, we start with the fact that every tree has either a center edge or a center vertex which is fixed by every automorphism of the tree. The center is obtained by removing every leaf and its incident edge successively until only a vertex or edge remains.

To prove the dissymmetry theorem, we start with the fact that every tree has either a center edge or a center vertex which is fixed by every automorphism of the tree. The center is obtained by removing every leaf and its incident edge successively until only a vertex or edge remains.

To prove the dissymmetry theorem, we start with the fact that every tree has either a center edge or a center vertex which is fixed by every automorphism of the tree. The center is obtained by removing every leaf and its incident edge successively until only a vertex or edge remains.

To prove the dissymmetry theorem, we start with the fact that every tree has either a center edge or a center vertex which is fixed by every automorphism of the tree. The center is obtained by removing every leaf and its incident edge successively until only a vertex or edge remains.



Now suppose that T^w is a T tree rooted at a vertex or edge w. If w is the center of T then $\phi(T^w)$ is just T. Otherwise, there is a unique path from w to the center. Then $\phi(T^w)$ is T rooted at both w and the next vertex or edge on this path (which might be the center).

A corollary of the dissymmetry theorem allows us to count (unrooted) unlabeled trees:

Let $\mathscr{T}$ be a class of trees closed under isomorphism. Let $T(x)$ be the generating function for unlabeled trees of $\mathscr{T}$. Let $T^{\bullet}(x)$, $T^{-}(x)$, and $T^{\bullet -}(x)$ be the generating functions for unlabeled trees in $\mathscr{T}$ rooted at a vertex, at an edge, or at a vertex and incident edge. Then

$$T(x) + T^{\bullet -}(x) = T^{\bullet}(x) + T^{-}(x),$$

so

$$T(x) = T^{\bullet}(x) + T^{-}(x) - T^{\bullet -}(x).$$

For the case in which $\mathscr{T}$ is the class of all trees, we saw that $T^{\bullet}(x) = R(x)$ where $R(x) = xh[R(x)]$. We need to compute $T^{-}(x)$ and $T^{\bullet -}(x)$.

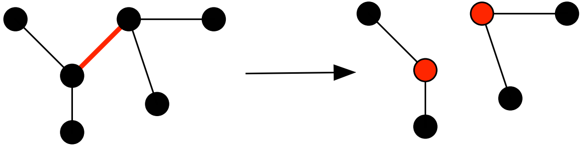
In fact, we have

$$T^{-}(x) = h_2[R(x)] = \frac{1}{2}\left(R(x)^2 + R(x^2)\right)$$

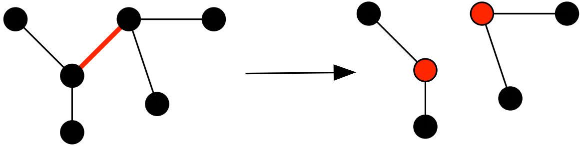
and

$$T^{\bullet -}(x) = R(x)^2.$$

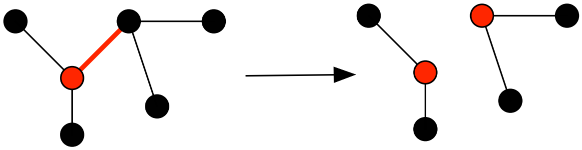
For $T^{-}(x) = h_2[R(x)]$, we have a bijection from trees rooted at an edge to unordered pairs of rooted trees:



For $T^{-}(x) = h_2[R(x)]$, we have a bijection from trees rooted at an edge to unordered pairs of rooted trees:



Similarly, for $T^{\bullet -}(x) = R(x)^2$, we have a bijection from trees rooted at a vertex and incident edge to ordered pairs of rooted trees:



Putting everything together, we get

$$\begin{aligned} T(x) &= R(x) + h_2[R(x)] - R(x)^2 \\ &= R(x) + \frac{1}{2}\Big(R(x)^2 + R(x^2)\Big) - R(x)^2 \\ &= R(x) - \frac{1}{2}\Big(R(x)^2 - R(x^2)\Big) \end{aligned}$$

where

$$R(x) = xh[R(x)] = x\exp\left(\sum_{k=1}^{\infty} \frac{R(x^k)}{k}\right)$$

and we can easily compute

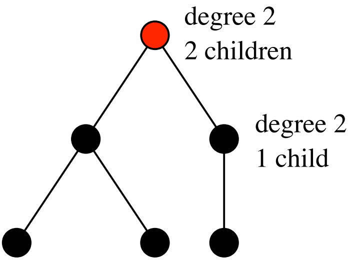
$$\begin{aligned} T(x) = x + x^2 + 3x^3 + 6x^4 + 15x^5 + 34x^6 + 85x^7 + 207x^8 + 525x^9 \\ + 1332x^{10} + 3449x^{11} + 8981x^{12} + 23671x^{13} + 62787x^{14} \\ + 167881x^{15} + 451442x^{16} + 1221065x^{17} + 3318451x^{18} \\ + 9059397x^{19} + 24829391x^{20} + 68299159x^{21} + \cdots \end{aligned}$$

Counting homeomorphically irreducible trees

We take the same approach with homeomorphically irreducible trees. We need to count homeomorphically irreducible trees rooted at a vertex, at an edge, and at a vertex and incident edge.

Homeomorphically irreducible trees rooted a vertex

Let’s compare the degrees of vertices with the number of children in rooted trees:



We need to count rooted trees in which the root does not have exactly two children, and no non-root vertex has exactly one child.

We need to count rooted trees in which the root does not have exactly two children, and no non-root vertex has exactly one child.

To do this we first count rooted trees in which no vertex has one child. Let $S(x)$ be the generating function for these trees. Then

$$S(x) = x \sum_{n \neq 1} h_n[S(x)] = x\big(h[S(x)] - S(x)\big)$$

We need to count rooted trees in which the root does not have exactly two children, and no non-root vertex has exactly one child.

To do this we first count rooted trees in which no vertex has one child. Let $S(x)$ be the generating function for these trees. Then

$$S(x) = x \sum_{n \neq 1} h_n[S(x)] = x\big(h[S(x)] - S(x)\big)$$

and the generating function for homeomorphically irreducible trees rooted at a vertex is

$$\begin{aligned} T^\bullet(x) &= x \sum_{n \neq 2} h_n[S(x)] = x\big(h[S(x)] - h_2[S(x)]\big). \\ &= (1+x)S(x) - xh_2[S(x)] \end{aligned}$$

The generating function for homeomorphically irreducible trees rooted at an edge is

$$T^{-}(x) = h_2[S(x)]$$

and the generating function for homeomorphically irreducible trees rooted a vertex and incident edge is

$$T^{\bullet -}(x) = S(x)^2.$$

So our final result is that the generating function for unrooted homeomorphically irreducible trees is

$$(1+x)S(x) + (1-x)h_2[S(x)] - S(x)^2$$

where $S(x)$ satisfies

$$S(x) = x\big(h[S(x)] - S(x)\big).$$

We can then compute as many terms as we want:

$$
\begin{aligned}
T(x) = x + x^2 + x^4 + x^5 + 2x^6 + 2x^7 + 4x^8 + 5x^9 + 10x^{10} \\
+ 14x^{11} + 26x^{12} + 42x^{13} + 78x^{14} + 132x^{15} + 249x^{16} \\
+ 445x^{17} + 842x^{18} + 1561x^{19} + 2988x^{20} + 5671x^{21} + \cdots
\end{aligned}
$$

We can then compute as many terms as we want:

$$T(x) = x + x^2 + x^4 + x^5 + 2x^6 + 2x^7 + 4x^8 + 5x^9 + 10x^{10} + 14x^{11} + 26x^{12} + 42x^{13} + 78x^{14} + 132x^{15} + 249x^{16} + 445x^{17} + 842x^{18} + 1561x^{19} + 2988x^{20} + 5671x^{21} + \cdots$$

The number of homeomorphically irreducible trees with 100 vertices is 7611990566708854733349833156.