Combinatorics of Phylogenetic Networks

Michael Fuchs

Department of Mathematical Sciences
Chengchi University

July 10th, 2023

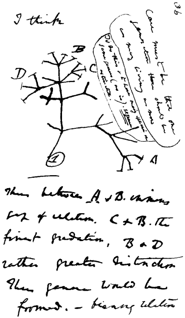
Evolutionary Biology

Charles Darwin
(1809-1882)

Evolutionary Biology

Charles Darwin
(1809-1882)

First notebook on Transmutation of Species (1837)

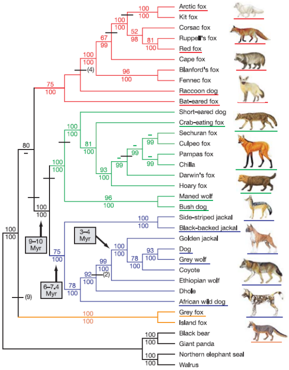


What is a Phylogenetic Tree?

X . . . a finite set.

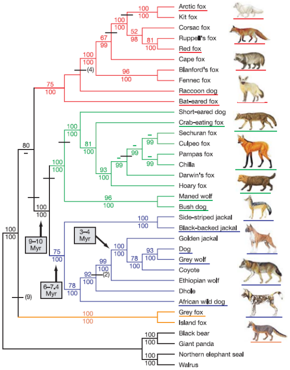
What is a Phylogenetic Tree?

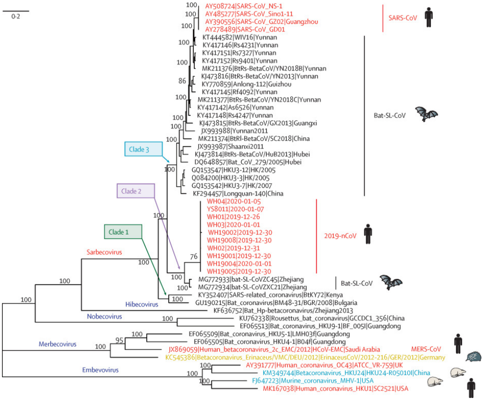
X . . . a finite set.



What is a Phylogenetic Tree?

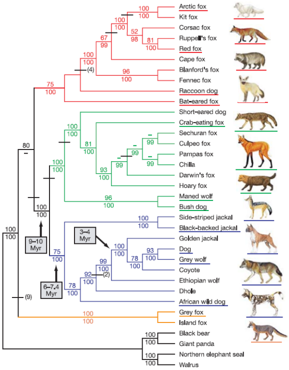
X . . . a finite set.

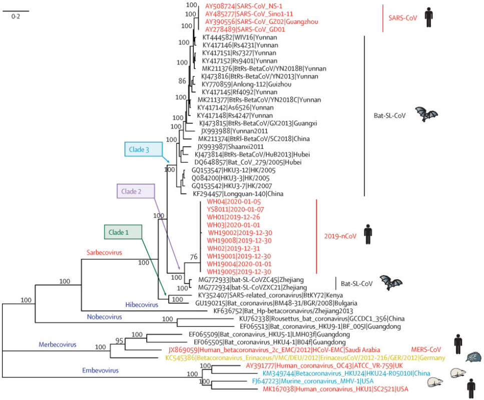




What is a Phylogenetic Tree?

X . . . a finite set.





A phylogenetic tree is a rooted, non-plane, binary tree with leaves labeled by X.

Counting Phylogenetic Trees

$\mathrm{T}_n \ldots \#$ of phylogenetic trees with n labeled leaves.

Counting Phylogenetic Trees

T_n ... # of phylogenetic trees with n labeled leaves.

Pick a phylogenetic tree with $n-1$ labeled leaves.

Counting Phylogenetic Trees

$\mathrm{T}_n \ldots \#$ of phylogenetic trees with n labeled leaves.

Pick a phylogenetic tree with $n-1$ labeled leaves.

Do the following:

Counting Phylogenetic Trees

$\mathrm{T}_n \ldots \#$ of phylogenetic trees with n labeled leaves.

Pick a phylogenetic tree with $n-1$ labeled leaves.

Do the following:

(i) either attach a pendant edge with a node labeled by n to any edge,

Counting Phylogenetic Trees

$\mathrm{T}_n \ldots \#$ of phylogenetic trees with n labeled leaves.

Pick a phylogenetic tree with $n-1$ labeled leaves.

Do the following:

(i) either attach a pendant edge with a node labeled by n to any edge,

(ii) or create a new root with one child the tree and the other the node labeled by n.

Counting Phylogenetic Trees

$\mathrm{T}_n \ldots \#$ of phylogenetic trees with n labeled leaves.

Pick a phylogenetic tree with $n-1$ labeled leaves.

Do the following:

(i) either attach a pendant edge with a node labeled by n to any edge,

(ii) or create a new root with one child the tree and the other the node labeled by n.

This produces all trees of n labeled leaves.

Counting Phylogenetic Trees

T_n ... # of phylogenetic trees with n labeled leaves.

Pick a phylogenetic tree with $n-1$ labeled leaves.

Do the following:

(i) either attach a pendant edge with a node labeled by n to any edge,

(ii) or create a new root with one child the tree and the other the node labeled by n.

This produces all trees of n labeled leaves.

Thus,

$$\mathrm{T}_n = (2n-3)\mathrm{T}_{n-1}$$

and by iteration

$$\mathrm{T}_n = (2n-3)!!.$$

What is a Phylogenetic Network?

X ... a finite set.

What is a Phylogenetic Network?

X ... a finite set.

Definition

A phylogenetic network is a rooted DAG which has the following nodes:

What is a Phylogenetic Network?

X . . . a finite set.

Definition

A phylogenetic network is a rooted DAG which has the following nodes:

(a) *root: in-degree* 0 *and out-degree* 1*;*

What is a Phylogenetic Network?

X . . . a finite set.

Definition

A phylogenetic network is a rooted DAG which has the following nodes:

(a) *root: in-degree* 0 *and out-degree* 1*;*

(b) *leaves: in-degree* 1 *and out-degree* 0*; bijectively labeled by* X*;*

What is a Phylogenetic Network?

X . . . a finite set.

Definition

A phylogenetic network is a rooted DAG which has the following nodes:

(a) *root: in-degree 0 and out-degree 1;*

(b) *leaves: in-degree 1 and out-degree 0; bijectively labeled by X;*

(c) *all other nodes have either out-degree 2 and in-degree 1 (tree nodes) or out-degree 1 and in-degree 2 (reticulation nodes).*

What is a Phylogenetic Network?

X . . . a finite set.

Definition

A phylogenetic network is a rooted DAG which has the following nodes:

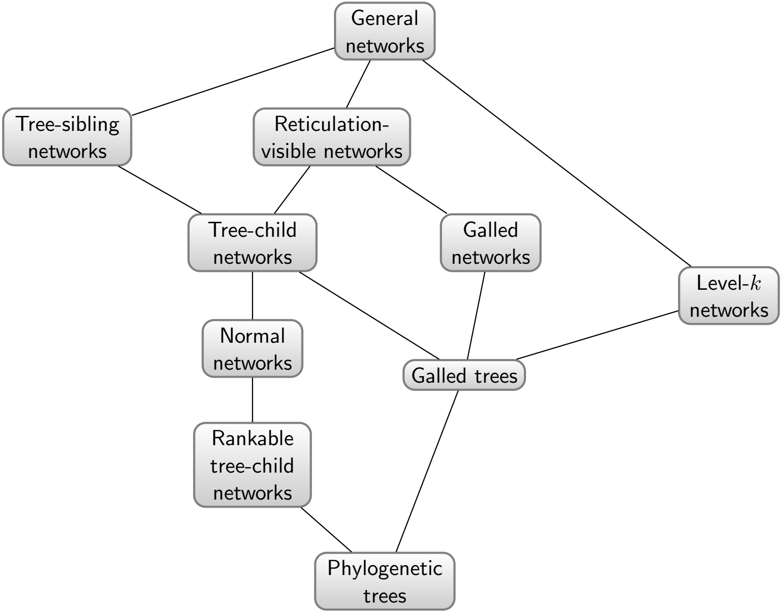
(a) *root: in-degree 0 and out-degree 1;*

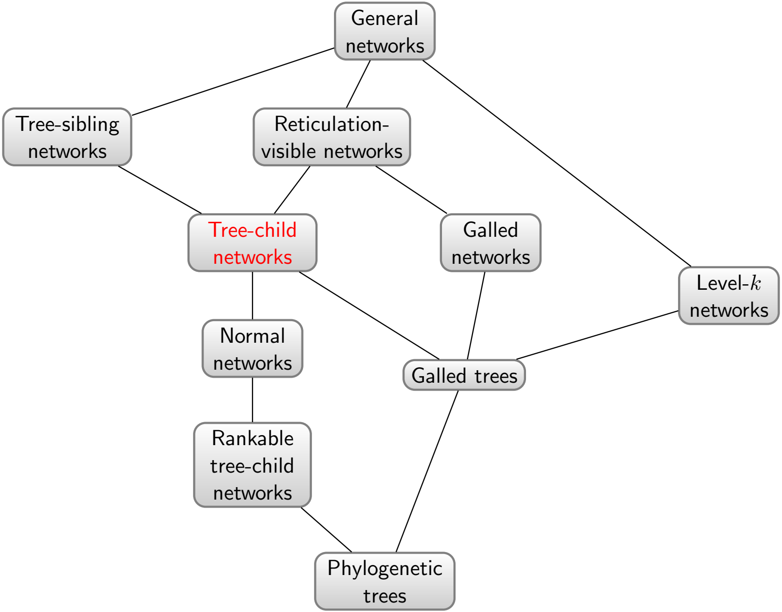
(b) *leaves: in-degree 1 and out-degree 0; bijectively labeled by X;*

(c) *all other nodes have either out-degree 2 and in-degree 1 (tree nodes) or out-degree 1 and in-degree 2 (reticulation nodes).*

Phylogenetic networks have become increasingly popular in recent decades.

They are used to model reticulate evolution which contains reticulation events such as lateral gene transfer or hybridization.


General networks
Tree-sibling networks
Reticulation-visible networks
Tree-child networks
Galled networks
Level-k networks
Normal networks
Galled trees
Rankable tree-child networks
Phylogenetic trees


General networks
Tree-sibling networks
Reticulation-visible networks
Tree-child networks
Galled networks
Level-k networks
Normal networks
Galled trees
Rankable tree-child networks
Phylogenetic trees

TC-Networks

Definition

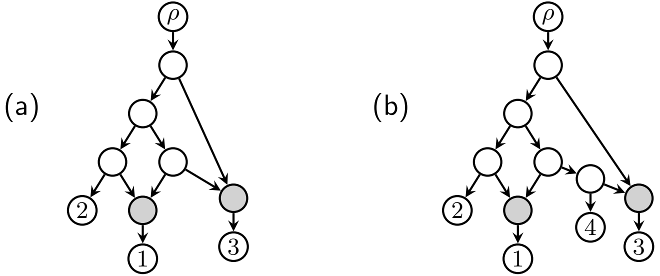
A phylogenetic network is called tree-child network if every non-leaf node has at least one child which is not a reticulation node.

TC-Networks

Definition

A phylogenetic network is called *tree-child network* *if every non-leaf node has at least one child which is not a reticulation node.*

Examples:



TC-Networks

Definition

A phylogenetic network is called tree-child network if every non-leaf node has at least one child which is not a reticulation node.

Examples:

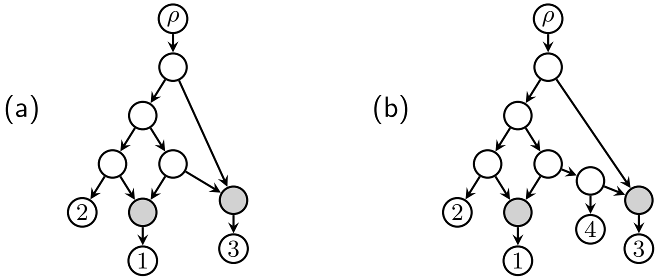


Figure: (a) is not a tc-network whereas (b) is a tc-network.

Method of Component Graphs

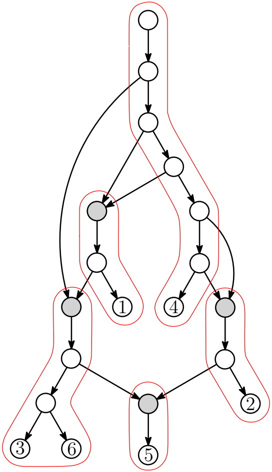
Method of Component Graphs

Cardona & Zhang (JCSS; 2020) used component graphs:

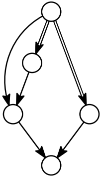
Method of Component Graphs

Cardona & Zhang (JCSS; 2020) used component graphs:

(a)



(b)



Counting TC-Networks

k_m ... # of component graphs with m nodes.

Proposition

k_m satisfies $k_m = \sum_{s=1}^{m-1} k_{m,s}$ where $k_{1,1} = 1$ and

$$k_{m,s} = \sum_{1 \leq t \leq m-1-s} \binom{m}{s} \sum_{0 \leq \ell \leq t} (-1)^{\ell} \binom{t}{\ell} \binom{m-1-s-\ell+d}{d} k_{m-s,t}.$$

Counting TC-Networks

k_m ... # of component graphs with m nodes.

Proposition

k_m satisfies $k_m = \sum_{s=1}^{m-1} k_{m,s}$ where $k_{1,1} = 1$ and

$$k_{m,s} = \sum_{1\leq t\leq m-1-s} \binom{m}{s} \sum_{0\leq \ell\leq t} (-1)^{\ell} \binom{t}{\ell} \binom{m-1-s-\ell+d}{d} k_{m-s,t}.$$

$\mathrm{TC}_{n,k}$ ... # of tc-networks with n leaves and k reticulation nodes.

Theorem (Cardona & Zhang; 2020)

$$\mathrm{TC}_{n,k} = \frac{1}{2^{n-1-k}} \sum_{\{B_j\}_{j=1}^{k+1}} \sum_{G\in\mathcal{K}_{k+1}} \prod_{j=1}^{k+1} \frac{(2b_j+g_j-2)}{(b_j-1)! \prod_{\ell=1}^{k+1}(g_{j,\ell})!}.$$

$\mathrm{TC}_{n,k}$ for small n, k

$\mathrm{TC}_{n,k}$ for small n, k

Lemma

In any tc-network: $k \leq n - 1$.

$\mathrm{TC}_{n,k}$ for small n, k

Lemma

In any tc-network: $k \leq n - 1$.

Cardona & Zhang:

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

$\mathrm{TC}_{n,k}$ for small n, k

Lemma

In any tc-network: $k \leq n-1$.

Cardona & Zhang:

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Computation becomes more and more cumbersome because the number of component graphs increases rapidly!

Formulas for small k

Theorem (Cardona & Zhang; 2020)

We have,

$$\mathrm{TC}_{n,1} = \frac{n!(2n)!}{2^n n!} - 2^{n-1} n!.$$

and

$$\mathrm{TC}_{n,2} = \frac{n!}{2^n} \sum_{j=1}^{n-2} \binom{2j}{j} \binom{2n-2j}{n-j} \frac{j(2j+1)(2n-j-1)}{2n-2j-1} + n(n-1)n!2^{n-3} - \frac{(2n-1)!n}{3 \cdot 2^{n-1}(n-2)!}$$

.

Formulas for small k

Theorem (Cardona & Zhang; 2020)

We have,

$$\mathrm{TC}_{n,1} = \frac{n!(2n)!}{2^n n!} - 2^{n-1} n!.$$

and

$$\begin{aligned}\mathrm{TC}_{n,2} &= \frac{n!}{2^n} \sum_{j=1}^{n-2} \binom{2j}{j} \binom{2n-2j}{n-j} \frac{j(2j+1)(2n-j-1)}{2n-2j-1} \\ &\quad + n(n-1)n!2^{n-3} - \frac{(2n-1)!n}{3 \cdot 2^{n-1}(n-2)!} \\ &= n(n-1)\left(\frac{3n+2}{3}(2n-1)!! - (2n)!!\right).\end{aligned}$$

Formulas for small k

Theorem (Cardona & Zhang; 2020)

We have,

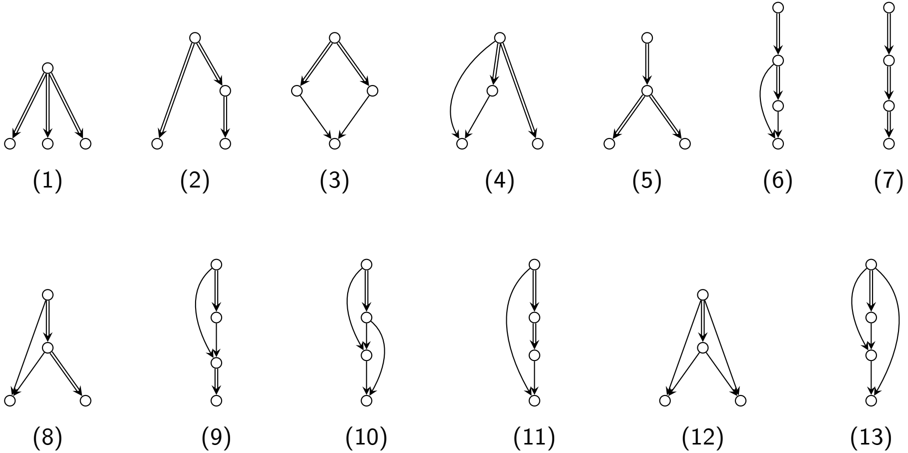
$$\mathrm{TC}_{n,1} = \frac{n!(2n)!}{2^n n!} - 2^{n-1} n!.$$

and

$$\mathrm{TC}_{n,2} = \frac{n!}{2^n} \sum_{j=1}^{n-2} \binom{2j}{j} \binom{2n-2j}{n-j} \frac{j(2j+1)(2n-j-1)}{2n-2j-1}$$
$$+ n(n-1)n!2^{n-3} - \frac{(2n-1)!n}{3 \cdot 2^{n-1}(n-2)!}$$
$$= n(n-1) \left(\frac{3n+2}{3}(2n-1)!! - (2n)!! \right).$$

En-Yu Huang (2022) derived a formula for $k = 3$.

Component Graphs for $k = 3$



Asymptotics of TC-Networks with fixed k

Proposition

Let $S_{n,k}$ be the number of tc-networks arising from the star-component graph. Then,

$$S_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!}\left(\frac{2}{e}\right)^{n} n^{n+2k-1}.$$

Asymptotics of TC-Networks with fixed k

Proposition

Let $S_{n,k}$ be the number of tc-networks arising from the star-component graph. Then,

$$S_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!}\left(\frac{2}{e}\right)^n n^{n+2k-1}.$$

In fact, as $n \to \infty$, $S_{n,k} \sim \mathrm{TC}_{n,k}$.

Asymptotics of TC-Networks with fixed k

Proposition

Let $S_{n,k}$ be the number of tc-networks arising from the star-component graph. Then,

$$S_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!}\left(\frac{2}{e}\right)^n n^{n+2k-1}.$$

In fact, as $n \to \infty$, $S_{n,k} \sim \mathrm{TC}_{n,k}$.

Theorem (F. & Huang & Yu; 2022)

As $n \to \infty$,

$$\mathrm{TC}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!}\left(\frac{2}{e}\right)^n n^{n+2k-1}.$$

Asymptotics of TC-Networks with fixed k

Proposition

Let $S_{n,k}$ be the number of tc-networks arising from the star-component graph. Then,

$$S_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!}\left(\frac{2}{e}\right)^n n^{n+2k-1}.$$

In fact, as $n \to \infty$, $S_{n,k} \sim \mathrm{TC}_{n,k}$.

Theorem (F. & Huang & Yu; 2022)

As $n \to \infty$,

$$\mathrm{TC}_{n,k} \sim \frac{2^{k-1}\sqrt{2}}{k!}\left(\frac{2}{e}\right)^n n^{n+2k-1}.$$

General TC-Networks (i)

$\mathrm{TC}_n \ldots \#$ of tc-networks with n leaves.

General TC-Networks (i)

TC_n ... # of tc-networks with n leaves.

Theorem (McDiarmid, Semple, Welsh; 2015)

For constants $0 < c_1 < c_2$,

$$(c_1 n)^{2n} \leq \mathrm{TC}_n \leq (c_2 n)^{2n}.$$

General TC-Networks (i)

TC_n ... # of tc-networks with n leaves.

Theorem (McDiarmid, Semple, Welsh; 2015)

For constants $0 < c_1 < c_2$,

$$(c_1 n)^{2n} \leq \mathrm{TC}_n \leq (c_2 n)^{2n}.$$

Question: what is the exponential growth rate?

General TC-Networks (i)

$\mathrm{TC}_n \ldots \#$ of tc-networks with n leaves.

Theorem (McDiarmid, Semple, Welsh; 2015)

For constants $0 < c_1 < c_2$,

$$(c_1 n)^{2n} \leq \mathrm{TC}_n \leq (c_2 n)^{2n}.$$

Question: what is the exponential growth rate?

McDiarmid & Semple & Welsh (2015) also proved stochastic results.

General TC-Networks (i)

$\mathrm{TC}_n \ldots \#$ of tc-networks with n leaves.

Theorem (McDiarmid, Semple, Welsh; 2015)

For constants $0 < c_1 < c_2$,

$$(c_1 n)^{2n} \le \mathrm{TC}_n \le (c_2 n)^{2n}.$$

Question: what is the exponential growth rate?

McDiarmid & Semple & Welsh (2015) also proved stochastic results.

Theorem (McDiarmid, Semple, Welsh; 2015)

(a) *$\#$ of reticulation nodes $\sim n$ for almost all tc-networks;*

(b) *The number of cherries is $o(n)$ for almost all tc-networks.*

General TC-Networks (ii)

Cardona & Zhang:

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

General TC-Networks (ii)

Cardona & Zhang:

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Note that $\mathrm{TC}_{n,k}$ is rapidly increasing from $0 \leq k \leq n-1$!

General TC-Networks (ii)

Cardona & Zhang:

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Note that $\mathrm{TC}_{n,k}$ is rapidly increasing from $0 \leq k \leq n-1$!

General TC-Networks (ii)

Cardona & Zhang:

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Note that $\mathrm{TC}_{n,k}$ is rapidly increasing from $0 \le k \le n-1$!

General TC-Networks (ii)

Cardona & Zhang:

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Note that $\mathrm{TC}_{n,k}$ is rapidly increasing from $0 \le k \le n-1$!

Guess:

$$\mathrm{TC}_n = \sum_{k=0}^{n-1} \mathrm{TC}_{n,k} = \Theta(\mathrm{TC}_{n,n-1}).$$

General TC-Networks (ii)

Cardona & Zhang:

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Note that $\mathrm{TC}_{n,k}$ is rapidly increasing from $0 \le k \le n-1$!

Guess:

$$\mathrm{TC}_n = \sum_{k=0}^{n-1} \mathrm{TC}_{n,k} = \Theta(\mathrm{TC}_{n,n-1}).$$

Searching in the OEIS

We have,

$$\{\mathrm{TC}_{n,n-1}\} = \{2, 42, 2544, 309000, 63068400, 19474761600, \dots\}.$$

Searching in the OEIS

We have,

$$\{\mathrm{TC}_{n,n-1}\} = \{2, 42, 2544, 309000, 63068400, 19474761600, \ldots\}.$$

This sequence is not in OEIS!

Searching in the OEIS

We have,

$$\{\mathrm{TC}_{n,n-1}\} = \{2, 42, 2544, 309000, 63068400, 19474761600, \ldots\}.$$

This sequence is not in OEIS!

We have,

$$\left\{\frac{\mathrm{TC}_{n,n-1}}{n!}\right\} = \{1, 7, 106, 2575, 87595, 3864040, \ldots\}.$$

Searching in the OEIS

We have,

$$\{\mathrm{TC}_{n,n-1}\} = \{2, 42, 2544, 309000, 63068400, 19474761600, \ldots\}.$$

This sequence is not in OEIS!

We have,

$$\left\{\frac{\mathrm{TC}_{n,n-1}}{n!}\right\} = \{1, 7, 106, 2575, 87595, 3864040, \ldots\}.$$

This sequence is in OEIS: A213863!

Searching in the OEIS

We have,

$$\{\mathrm{TC}_{n,n-1}\} = \{2, 42, 2544, 309000, 63068400, 19474761600, \ldots\}.$$

This sequence is not in OEIS!

We have,

$$\left\{\frac{\mathrm{TC}_{n,n-1}}{n!}\right\} = \{1, 7, 106, 2575, 87595, 3864040, \ldots\}.$$

This sequence is in OEIS: A213863!

A213863 was submitted on June 23rd, 2012 by Alois P. Heinz who gave its first 17 terms and a (brute-force) Maple program to compute them; tc-networks are not mentioned in his entry.

A Counting Sequence of Words

Definition (OEIS; A213863)

Denote by a_n the number of words on letters $\{\omega_1, \dots, \omega_n\}$ so that

(i) *each letter occurs exactly 3 times;*

(ii) *ω_i has either not occurred or it has occurred at least as often as ω_j with $j > i$.*

A Counting Sequence of Words

Definition (OEIS; A213863)

Denote by a_n the number of words on letters $\{\omega_1, \dots, \omega_n\}$ so that

(i) *each letter occurs exactly 3 times;*

(ii) *ω_i has either not occurred or it has occurred at least as often as ω_j with $j > i$.*

For example, $a_2 = 7$ because

$$aaabbb, aababb, aabbab, abaabb, ababab, baaabb, baabab.$$

A Counting Sequence of Words

Definition (OEIS; A213863)

Denote by a_n the number of words on letters $\{\omega_1, \dots, \omega_n\}$ so that

(i) *each letter occurs exactly 3 times;*

(ii) *ω_i has either not occurred or it has occurred at least as often as ω_j with $j > i$.*

For example, $a_2 = 7$ because

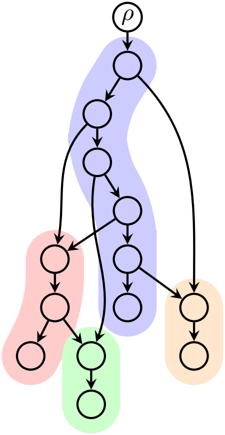
$$aaabbb, aababb, aabbab, abaabb, ababab, baaabb, baabab.$$

Proposition

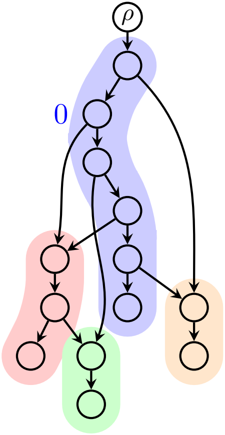
We have,

$$\frac{\mathrm{TC}_{n,n-1}}{n!} = a_{n-1}.$$

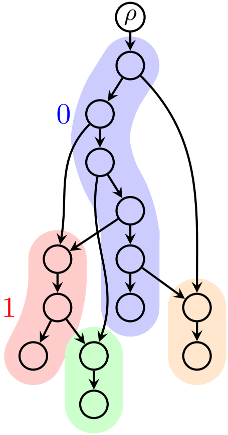
Bijective Proof



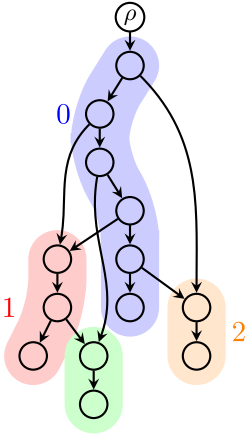
Bijective Proof



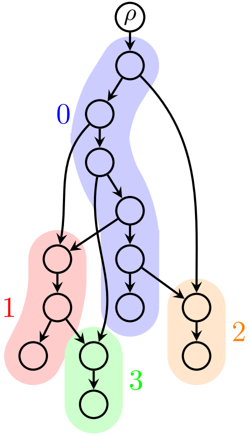
Bijective Proof



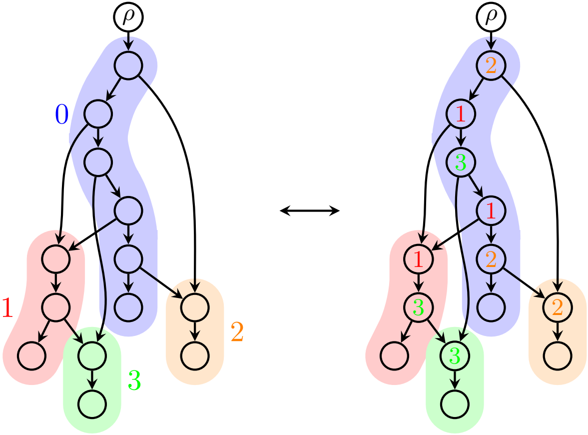
Bijective Proof



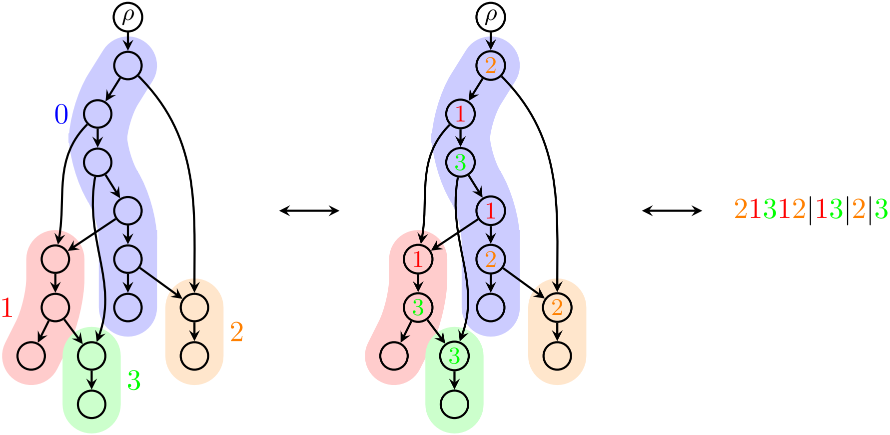
Bijective Proof



Bijective Proof



Bijective Proof



Asymptotics of General TC-Networks

Define

$$b_{n,m} = (2n + m - 2) \sum_{j=1}^{m} b_{n-1,j}$$

Then, $a_n = \sum_{m \geq 1} b_{n,m}$.

Asymptotics of General TC-Networks

Define

$$b_{n,m} = (2n+m-2)\sum_{j=1}^{m} b_{n-1,j}$$

Then, $a_n = \sum_{m\geq 1} b_{n,m}$.

From this, by a recent method of Elvey Price, Fang, Wallner (2021):

Asymptotics of General TC-Networks

Define

$$b_{n,m} = (2n+m-2)\sum_{j=1}^{m} b_{n-1,j}$$

Then, $a_n = \sum_{m\geq 1} b_{n,m}$.

From this, by a recent method of Elvey Price, Fang, Wallner (2021):

Theorem (F. & Yu & Zhang; 2021)

We have,

$$\mathrm{TC}_n = \Theta\left(n^{-2/3}e^{a_1(3n)^{1/3}}\left(\frac{12}{e^2}\right)^n n^{2n}\right),$$

where a_1 is the largest root of the Airy function of first order.

Asymptotics of General TC-Networks

Define

$$b_{n,m} = (2n+m-2)\sum_{j=1}^{m} b_{n-1,j}$$

Then, $a_n = \sum_{m\geq 1} b_{n,m}$.

From this, by a recent method of Elvey Price, Fang, Wallner (2021):

Theorem (F. & Yu & Zhang; 2021)

We have,

$$\mathrm{TC}_n = \Theta\left(n^{-2/3} e^{a_1 (3n)^{1/3}} \left(\frac{12}{e^2}\right)^n n^{2n}\right),$$

where a_1 is the largest root of the Airy function of first order.

Exact Enumeration of General TC-Networks

Definition

Denote by $w_{n,k}$ the number of words on letters $\{\omega_1, \ldots, \omega_n\}$ so that

(i) *k letters occur 3 times; $n-k$ letters occur 2 times, where the 0-th, 1-st, 2-nd occurrence counts as the 1-st, 2-nd, 3-rd.*

(ii) *ω_i has either not occurred or it has occurred at least as often as ω_j with $j > i$.*

Exact Enumeration of General TC-Networks

Definition

Denote by $w_{n,k}$ *the number of words on letters* $\{\omega_1, \dots, \omega_n\}$ *so that*

(i) k *letters occur* 3 *times;* $n-k$ *letters occur* 2 *times, where the* 0*-th,* 1*-st,* 2*-nd occurrence counts as the* 1*-st,* 2*-nd,* 3*-rd.*

(ii) ω_i *has either not occurred or it has occurred at least as often as* ω_j *with* $j > i$*.*

For example, $w_{2,1} = 7$ because

$$abbab, babab, aabbb, ababb, baabb, aabab, aaabb.$$

Exact Enumeration of General TC-Networks

Definition

Denote by $w_{n,k}$ the number of words on letters $\{\omega_1, \dots, \omega_n\}$ so that

(i) *k letters occur 3 times; $n-k$ letters occur 2 times, where the 0-th, 1-st, 2-nd occurrence counts as the 1-st, 2-nd, 3-rd.*

(ii) *ω_i has either not occurred or it has occurred at least as often as ω_j with $j > i$.*

For example, $w_{2,1} = 7$ because

$$abbab, babab, aabbb, ababb, baabb, aabab, aaabb.$$

Theorem (Chang & F. & Liu & Wallner & Yu; 2022+)

We have,

$$\frac{\mathrm{TC}_{n,k}}{n!} = \frac{w_{n-1,k}}{2^{n-1-k}}.$$

Fast Computation of $\mathrm{TC}_{n,k}$

Define,

$$b_{n,k,m} = \sum_{j=1}^{m} b_{n-1,k,j} + (n+m+k-2) \sum_{j=1}^{m} b_{n-1,k-1,j}$$

Then, $\omega_{n,k} = \sum_{m\geq 1} b_{n,k,m}$.

Fast Computation of $\mathrm{TC}_{n,k}$

Define,

$$b_{n,k,m} = \sum_{j=1}^{m} b_{n-1,k,j} + (n+m+k-2) \sum_{j=1}^{m} b_{n-1,k-1,j}$$

Then, $\omega_{n,k} = \sum_{m\geq 1} b_{n,k,m}$.

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Fast Computation of $\mathrm{TC}_{n,k}$

Define,

$$b_{n,k,m} = \sum_{j=1}^{m} b_{n-1,k,j} + (n+m+k-2) \sum_{j=1}^{m} b_{n-1,k-1,j}$$

Then, $\omega_{n,k} = \sum_{m\geq 1} b_{n,k,m}$.

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Fast Computation of $\mathrm{TC}_{n,k}$

Define,

$$b_{n,k,m} = \sum_{j=1}^{m} b_{n-1,k,j} + (n+m+k-2)\sum_{j=1}^{m} b_{n-1,k-1,j}$$

Then, $\omega_{n,k} = \sum_{m\geq 1} b_{n,k,m}$.

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Fast Computation of $\mathrm{TC}_{n,k}$

Define,

$$b_{n,k,m} = \sum_{j=1}^{m} b_{n-1,k,j} + (n+m+k-2) \sum_{j=1}^{m} b_{n-1,k-1,j}$$

Then, $\omega_{n,k} = \sum_{m \geq 1} b_{n,k,m}$.

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Fast Computation of $\mathrm{TC}_{n,k}$

Define,

$$b_{n,k,m} = \sum_{j=1}^{m} b_{n-1,k,j} + (n+m+k-2) \sum_{j=1}^{m} b_{n-1,k-1,j}$$

Then, $\omega_{n,k} = \sum_{m\geq 1} b_{n,k,m}$.

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Fast Computation of $\mathrm{TC}_{n,k}$

Define,

$$b_{n,k,m} = \sum_{j=1}^{m} b_{n-1,k,j} + (n+m+k-2) \sum_{j=1}^{m} b_{n-1,k-1,j}$$

Then, $\omega_{n,k} = \sum_{m\geq 1} b_{n,k,m}$.

$k \setminus n$	2	3	4	5	6	7
1	2	21	228	2805	39330	623385
2		42	1272	30300	696600	16418430
3			2544	154500	6494400	241204950
4				309000	31534200	2068516800
5					63068400	9737380800
6						19474761600

Stochastic Results for TC-Networks

From the table, one can observe that:

$$\mathrm{TC}_{n,n-1-k} \approx \frac{1}{2^k k!}\mathrm{TC}_{n,n-1}$$

for k close to n.

Stochastic Results for TC-Networks

From the table, one can observe that:

$$\mathrm{TC}_{n,n-1-k} \approx \frac{1}{2^k k!}\mathrm{TC}_{n,n-1}$$

for k close to n.

Theorem (Chang, F. & Liu & Wallner & Yu; 2023+)

(a) *We have,*

$$n - 1 - \#\ \textit{of reticulation nodes} \xrightarrow{d} \textit{Poisson}(1/2).$$

Stochastic Results for TC-Networks

From the table, one can observe that:

$$\mathrm{TC}_{n,n-1-k} \approx \frac{1}{2^k k!}\mathrm{TC}_{n,n-1}$$

for k close to n.

Theorem (Chang, F. & Liu & Wallner & Yu; 2023+)

(a) *We have,*

$$n - 1 - \#\ \textit{of reticulation nodes} \xrightarrow{d} \textit{Poisson}(1/2).$$

(b) *We have,*

$$\mathbb{E}(\#\ \textit{of cherries}) = \mathcal{O}(1).$$

Multicombining TC-Networks

Definition

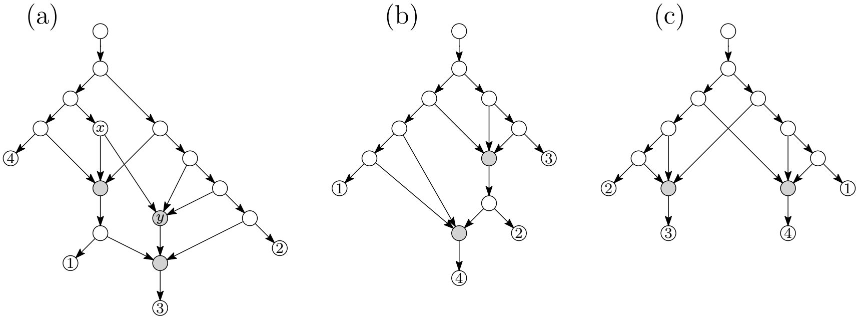
Let $d \geq 2$. A d-combining tree-child network is a tree-child network with each reticulation node having exactly d parents.

Multicombining TC-Networks

Definition

Let $d \geq 2$. A d-combining tree-child network is a tree-child network with each reticulation node having exactly d parents.

Examples: $d = 3$



Asymptotic Counting Result

Our encoding by words also works for d-combining tc-networks.

Asymptotic Counting Result

Our encoding by words also works for d-combining tc-networks.

$\mathrm{TC}_n^{[d]}$ ... # of d-combining networks with n leaves.

Asymptotic Counting Result

Our encoding by words also works for d-combining tc-networks.

$\mathrm{TC}_n^{[d]}$ ... # of d-combining networks with n leaves.

Following the same strategy as for $d = 2$ then gives the following result.

Asymptotic Counting Result

Our encoding by words also works for d-combining tc-networks.

$\mathrm{TC}_n^{[d]}$ ... # of d-combining networks with n leaves.

Following the same strategy as for $d = 2$ then gives the following result.

Theorem (Chang, F. & Liu & Wallner & Yu; 2023+)

We have,

$$\mathrm{TC}_n^{[d]} = \Theta\left((n!)^d\,\gamma(d)^n\,e^{3a_1\beta(d)n^{1/3}}n^{\alpha(d)}\right),$$

where a_1 is the largest root of the Airy function of the first kind and

$$\alpha(d) = -\frac{d(3d-1)}{2(d+1)}, \qquad \beta(d) = \left(\frac{d-1}{d+1}\right)^{2/3}, \qquad \gamma(d) = 4\frac{(d+1)^{d-1}}{(d-1)!}.$$

The Number of Reticulation Nodes

$\mathrm{TC}_{n,k}^{[d]}$ ... # of d-combining tc-networks with n leaves and k reticulation nodes.

The Number of Reticulation Nodes

$\mathrm{TC}_{n,k}^{[d]}$ ... # of d-combining tc-networks with n leaves and k reticulation nodes.

We have the bound:

$$\mathrm{TC}_{n,n-1-k}^{[d]} \leq \frac{1}{2^k k!} \mathrm{TC}_{n,n-1}^{[d]}$$

but for $d \geq 3$ it is not sharp anymore for k close to n!

The Number of Reticulation Nodes

$\mathrm{TC}_{n,k}^{[d]}$ ... # of d-combining tc-networks with n leaves and k reticulation nodes.

We have the bound:

$$\mathrm{TC}_{n,n-1-k}^{[d]} \leq \frac{1}{2^k k!}\mathrm{TC}_{n,n-1}^{[d]}$$

but for $d \geq 3$ it is not sharp anymore for k close to n!

Theorem (Chang & F. & Liu & Wallner & Yu; 2023+)

Let $d \geq 3$. The limit law of the number of reticulation nodes is degenerate. More precisely,

$$n - 1 - \text{\# of reticulation nodes} \xrightarrow{L_1} 0.$$

Fixed Number of Reticulation Nodes

The component method can also be extended to d-combining tc-networks.

Fixed Number of Reticulation Nodes

The component method can also be extended to d-combining tc-networks.

Theorem (Chang & F. & Liu & Wallner & Yu; 2023+)

We have,

$$\mathrm{TC}_{n,1}^{[3]} = \frac{n(2n+1)}{3}(2n-1)!! - n^2(2n-2)!!;$$

$$\mathrm{TC}_{n,2}^{[3]} = n(n-1)\left(\frac{70n^2+244n+177}{315}(2n+1)!! - \frac{16n+13}{48}(2n+2)!!\right).$$

In addition, as $n \to \infty$*,*

$$\mathrm{TC}_{n,k} \sim \frac{2^{dk-1}\sqrt{2}}{(d!)^k k!}\left(\frac{2}{e}\right)^n n^{n+dk-1}$$

for any fixed k*.*

Open Problems

- Can one obtain the first-order asymptotics? E.g., for $d = 2$, is there a constant c such that

$$\mathrm{TC}_n \sim c n^{-2/3} e^{a_1 (3n)^{1/3}} \left(\frac{12}{e^2}\right)^n n^{2n}?$$

Open Problems

- Can one obtain the first-order asymptotics? E.g., for $d = 2$, is there a constant c such that

$$\mathrm{TC}_n \sim c n^{-2/3} e^{a_1 (3n)^{1/3}} \left(\frac{12}{e^2}\right)^n n^{2n}?$$

Note that the Poisson limit law result implies that

$$\mathrm{TC}_n \sim \sqrt{e} \cdot \mathrm{TC}_{n,n-1}.$$

So, in order to prove the above claim, one needs to improve the method of Elvey Price, Fang, Wallner (2021).

Open Problems

- Can one obtain the first-order asymptotics? E.g., for $d = 2$, is there a constant c such that

$$\mathrm{TC}_n \sim cn^{-2/3}e^{a_1(3n)^{1/3}}\left(\frac{12}{e^2}\right)^n n^{2n}?$$

  Note that the Poisson limit law result implies that

$$\mathrm{TC}_n \sim \sqrt{e}\cdot \mathrm{TC}_{n,n-1}.$$

  So, in order to prove the above claim, one needs to improve the method of Elvey Price, Fang, Wallner (2021).

- Is it possible to derive limit laws for the number of cherries and more general patterns?

Open Problems

- Can one obtain the first-order asymptotics? E.g., for $d = 2$, is there a constant c such that

$$\mathrm{TC}_n \sim c n^{-2/3} e^{a_1 (3n)^{1/3}} \left(\frac{12}{e^2}\right)^n n^{2n}?$$

  Note that the Poisson limit law result implies that

$$\mathrm{TC}_n \sim \sqrt{e} \cdot \mathrm{TC}_{n,n-1}.$$

  So, in order to prove the above claim, one needs to improve the method of Elvey Price, Fang, Wallner (2021).

- Is it possible to derive limit laws for the number of cherries and more general patterns?

- How about results for height and Sackin index?

Some References

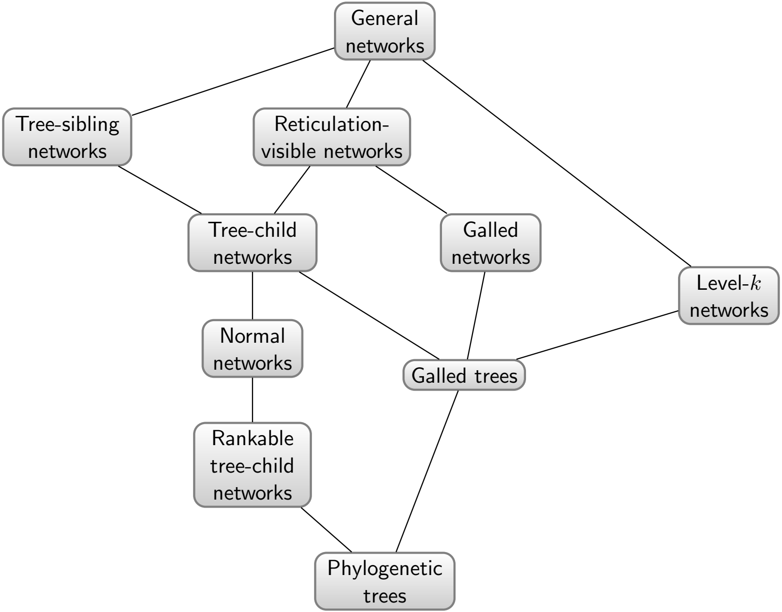
1. G. Cardona and L. Zhang (2020). Counting and enumerating tree-child networks and their subclasses, *J. Comput. System Sci.*, **114**, 84–104.

2. Y.-S. Chang, M. Fuchs, H. Liu, M. Wallner, G.-R. Yu. Enumerative and distributional results for d-combining tree-child networks, 48 pages, submitted.

3. M. Fuchs, E.-Y. Huang, G.-R. Yu (2022). Counting phylogenetic networks with few reticulation vertices: a second approach, *Discrete Appl. Math.*, **320**, 140–149.

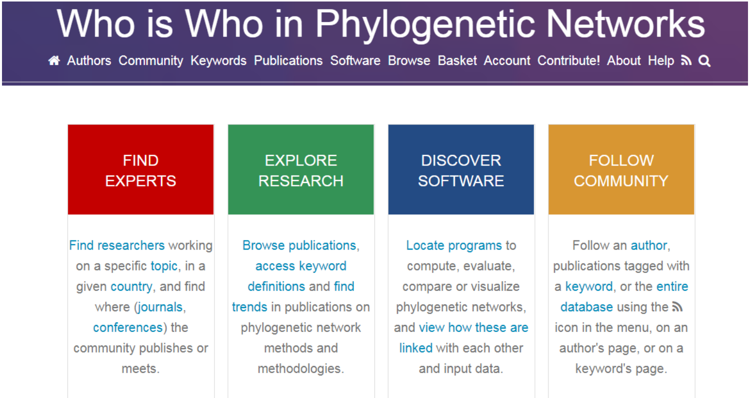
4. M. Fuchs, G.-R. Yu, L. Zhang (2021). On the asymptotic growth of the number of tree-child networks, *European J. Combin.*, **93**, 103278, 20 pages.

5. C. McDiarmid, C. Semple, D. Welsh (2015). Counting phylogenetic networks, *Ann. Comb.*, **19:1**, 205–224.


General networks
Tree-sibling networks
Reticulation-visible networks
Tree-child networks
Galled networks
Level-k networks
Normal networks
Galled trees
Rankable tree-child networks
Phylogenetic trees

There are more classes of phylogenetic networks:

http://phylnet.univ-mlv.fr/



Thanks for your attention!