Dyck-Łukasievicz trees: algebraic decomposition and random generation

Gilles Schaeffer
LIX, CNRS, Institut Polytechnique de Paris

based on joined work with Enrica Duchi,
IRIF, Université Paris Cité

GASCom 2022
Hommage à Jean-Guy Penaud
June 13, 2022, Varese

Summary of the talk

A few words about Jean-Guy Penaud

Bicolored binary trees and Dyck Łukasiewicz trees

A direct recursive approach and its complexity

Algebraic decompositions for marked trees

Complexity of random generation and a resampling trick

Conclusion

Some highlights in Jean-Guy's work

Dulucq-Penaud conjecture on Cori-Vauquelin trees for non-separable planar maps:

A non trivial characterization of the well labelled trees that correspond to non separable maps
(one of my first research interests, later rediscovered by Bouttier, Guitter, 2007)

A. del Lungo, F. del Ristoro and J.-G. Penaud's left ternary trees

See Enrica Duchi's talk

Bétréma-Penaud's proof from the book (quoting Doron Zeilberger)

The decomposition of pyramids of dominoes: all time favorite example of algebraic decomposition...

Viennot introduced pyramids of domino and obtained Motzkin like algebraic equations for their gf but the **direct interpretation** of these algebraic equations was given by Bétréma and Penaud.

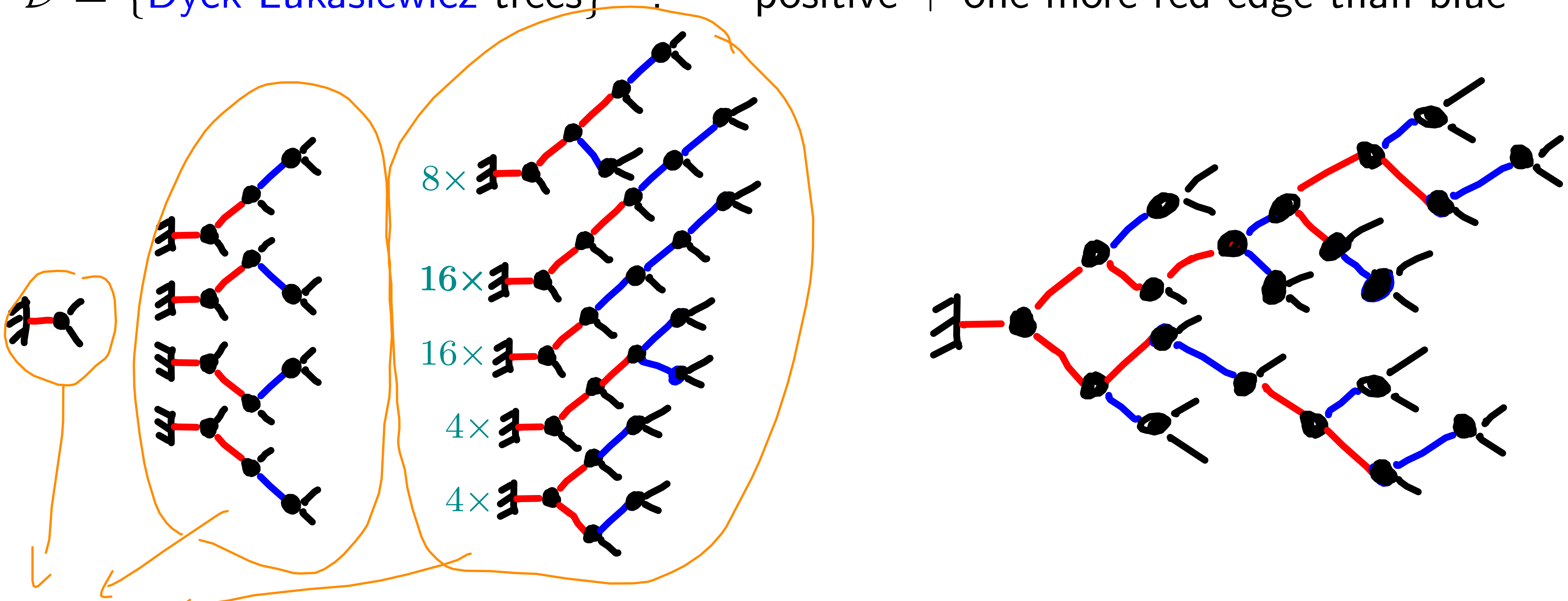
Bicolored binary trees and Dyck-Łukasiewicz trees

Dyck-Łukasiewicz trees

$\mathcal{B}$ = {blue/red binary trees} : planted binary tree with blue and red edges

$\mathcal{P}$ = {Positive bicolored trees} : no more red than blue in each planted subtree

$\mathcal{D}$ = {Dyck-Łukasiewicz trees} : positive + one more red edge than blue



1, 4, 48, 832, 17408, 408576, 10362880, 277954560, 7777026048, 224908017664

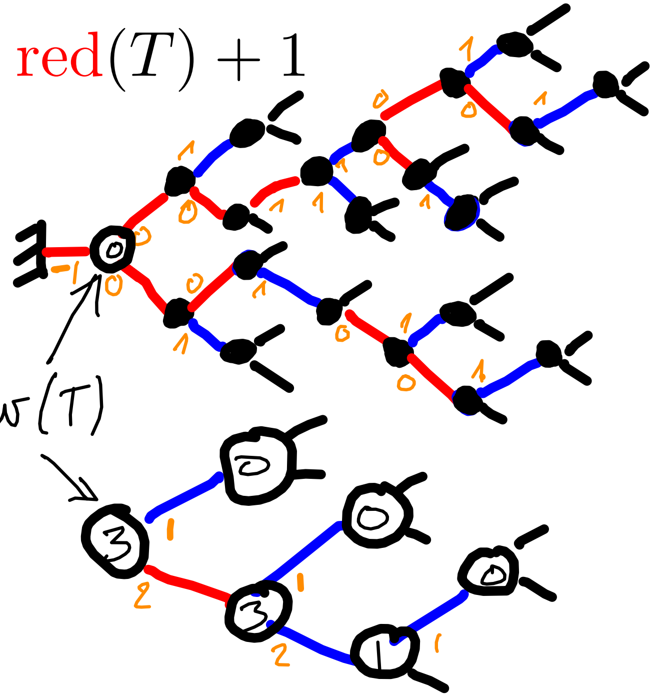
(fun game if you are tired of listening to talks: guess formula... you have 5 min before I give it)

A catalytic decomposition for positive bicolored trees

Let $F(u) \equiv F(u,t) = \sum_{T \in \mathcal{P}} u^{w(T)} t^{|T|}$, with $w(T) = \textcolor{blue}{\text{blue}}(T) - \textcolor{red}{\text{red}}(T) + 1$

so that $f \equiv f(t) = [u^0]F(u) = \sum_{T \in \mathcal{D}} t^{|T|}$ is the gf of Dyck trees

and more generally $F_m = [u^m]F(u)$ is the gf of positive tree with root vertex weight m.

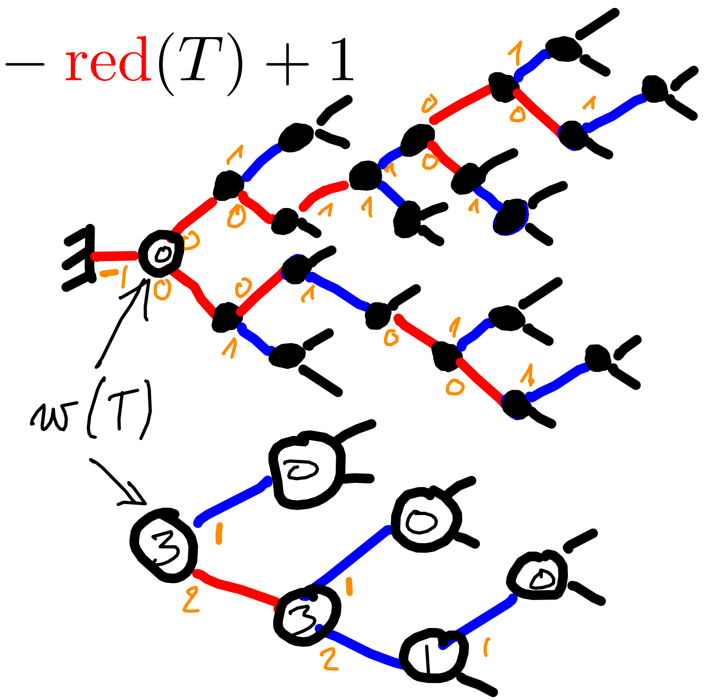


A catalytic decomposition for positive bicolored trees

Let $F(u) \equiv F(u,t) = \sum_{T \in \mathcal{P}} u^{w(T)} t^{|T|}$, with $w(T) = \text{blue}(T) - \text{red}(T) + 1$

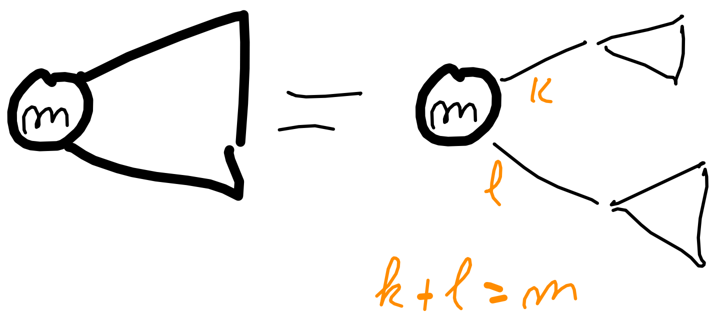
so that $f \equiv f(t) = [u^0]F(u) = \sum_{T \in \mathcal{D}} t^{|T|}$ is the gf of Dyck trees

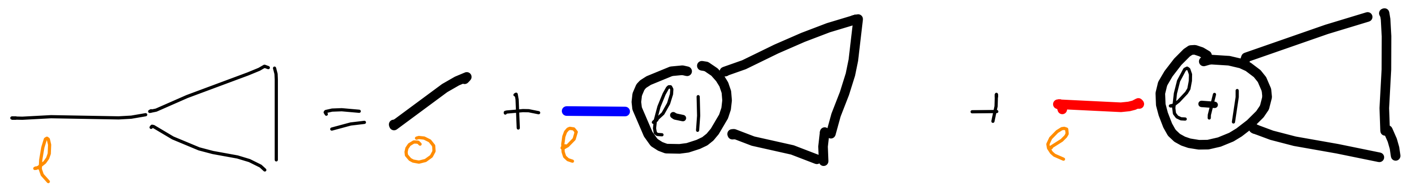
and more generally $F_m = [u^m]F(u)$ is the gf of positive tree with root vertex weight m.



Then:

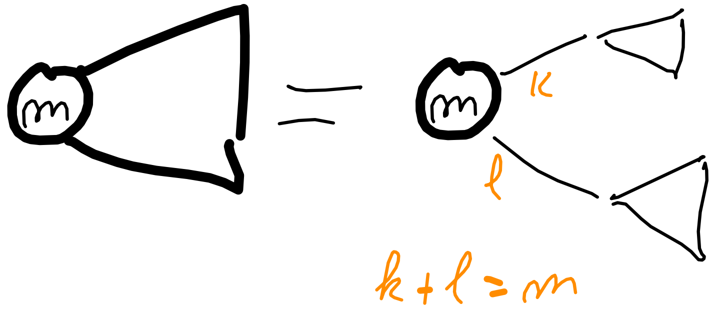
$$F(u) = tX(u)^2 \qquad \text{with} \qquad X(u) = 1 + u \cdot F(u) + \frac{F(u) - f}{u}$$

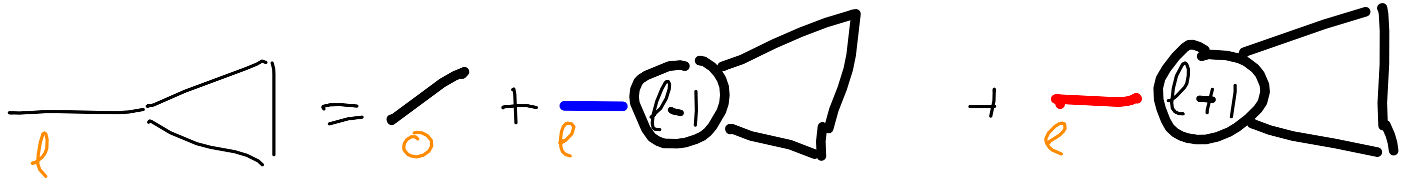




Random generation via catalytic decompositions

$$F(u) = tX(u)^2 \qquad \text{with} \qquad X(u) = 1 + u \cdot F(u) + \frac{F(u) - f}{u}$$





Amenable to a bivariate recursive approach (naive cubic complexity)

but not easily dealt with via Boltzmann due to the divided difference operator.

One variable / one function catalytic equations are easy

Bousquet-Mélou–Jehanne's trick gives an algebraic system

$\frac{\partial}{\partial u}$ applied to $$F(u) = t\left(1 + u\,F(u) + \frac{F(u)-f}{u}\right)^2$$

yields
$$\frac{\partial}{\partial u}F(u) = \frac{\partial}{\partial u}F(u)\cdot 2t\,\left(u+\tfrac{1}{u}\right)\left(1+u\,F(u)+\tfrac{F(u)-f}{u}\right)$$
$$+2\left(F(u)-\tfrac{1}{u}\,\tfrac{F(u)-f}{u}\right)\left(1+u\,F(u)+\tfrac{F(u)-f}{u}\right)$$

One variable / one function catalytic equations are easy

Bousquet-Mélou–Jehanne's trick gives an algebraic system

$\frac{\partial}{\partial u}$ applied to $$F(u) = t\left(1 + u\,F(u) + \frac{F(u) - f}{u}\right)^2$$

yields $$\frac{\partial}{\partial u}F(u) = \frac{\partial}{\partial u}F(u) \cdot 2t\,\left(u + \tfrac{1}{u}\right)\left(1 + u\,F(u) + \tfrac{F(u)-f}{u}\right)$$
$$+2\left(F(u) - \tfrac{1}{u}\,\tfrac{F(u)-f}{u}\right)\left(1 + u\,F(u) + \tfrac{F(u)-f}{u}\right)$$

Let $U \equiv U(t)$ be the unique fps s.t. $U = 2t\,(U^2+1)\left(1 + U\,F(U) + \frac{F(U)-f}{U}\right)$

cancels

One variable / one function catalytic equations are easy

Bousquet-Mélou–Jehanne's trick gives an algebraic system

$\frac{\partial}{\partial u}$ applied to $$F(u) = t \left(1 + u\,F(u) + \frac{F(u) - f}{u}\right)^2$$

yields $$\frac{\partial}{\partial u} F(u) = \frac{\partial}{\partial u} F(u) \cdot 2t \; \left(u + \tfrac{1}{u}\right) \left(1 + u\,F(u) + \tfrac{F(u) - f}{u}\right)$$
$$+2 \left(F(u) - \tfrac{1}{u} \; \tfrac{F(u) - f}{u}\right) \left(1 + u\,F(u) + \tfrac{F(u) - f}{u}\right)$$

cancels

Let $U \equiv U(t)$ be the unique fps s.t. $U = 2t \; (U^2 + 1) \left(1 + U\,F(U) + \tfrac{F(U) - f}{U}\right)$

U exists and has positive integer coeffs

One variable / one function catalytic equations are easy

Bousquet-Mélou–Jehanne's trick gives an algebraic system

$\frac{\partial}{\partial u}$ applied to $$F(u) = t\left(1 + u\,F(u) + \frac{F(u)-f}{u}\right)^2$$

yields $$\frac{\partial}{\partial u}F(u) = \frac{\partial}{\partial u}F(u)\cdot 2t\,\left(u+\tfrac{1}{u}\right)\left(1+u\,F(u)+\frac{F(u)-f}{u}\right)$$
$$+2\left(F(u)-\frac{1}{u}\,\frac{F(u)-f}{u}\right)\left(1+u\,F(u)+\frac{F(u)-f}{u}\right)$$

cancels

Let $U \equiv U(t)$ be the unique fps s.t. $U = 2t\,(U^2+1)\left(1+U\,F(U)+\frac{F(U)-f}{U}\right)$

U exists and has positive integer coeffs

then the series U, $V = F(U)$ and $W = \frac{F(U)-f}{U}$ satisfy the system:

$$\begin{cases} U &= 2t\,(U^2+1)\,(1+U\,V+W) \\ 0 &= U\,V - W \\ V &= t\,(1+U\,V+W)^2 \end{cases}$$

One variable / one function catalytic equations are easy

Bousquet-Mélou–Jehanne's trick gives an algebraic system

$\frac{\partial}{\partial u}$ applied to $\quad F(u) = t\left(1 + u\,F(u) + \dfrac{F(u)-f}{u}\right)^2$

yields $\quad \frac{\partial}{\partial u}F(u) = \frac{\partial}{\partial u}F(u) \cdot 2t\,\left(u + \frac{1}{u}\right)\left(1 + u\,F(u) + \frac{F(u)-f}{u}\right)$

$\quad +2\left(F(u) - \frac{1}{u}\,\frac{F(u)-f}{u}\right)\left(1 + u\,F(u) + \frac{F(u)-f}{u}\right)$

cancels

Let $U \equiv U(t)$ be the unique fps s.t. $U = 2t\,(U^2+1)\left(1 + U\,F(U) + \frac{F(U)-f}{U}\right)$

U exists and has positive integer coeffs

then the series U, $V = F(U)$ and $W = \frac{F(U)-f}{U}$ satisfy the system:

$$\begin{cases} U &= 2t\,(U^2+1)\,(1 + U\,V + W) \\ 0 &= U\,V - W \\ V &= t\,(1 + U\,V + W)^2 \end{cases}$$

and $\quad f = V - U \cdot W, \quad$ by definition of U, V, W.

One variable / one function catalytic equations are easy

Bousquet-Mélou–Jehanne's trick gives an algebraic system

$\frac{\partial}{\partial u}$ applied to $$F(u) = t\left(1 + u\,F(u) + \frac{F(u)-f}{u}\right)^2$$

yields $$\frac{\partial}{\partial u}F(u) = \frac{\partial}{\partial u}F(u)\cdot 2t\,\left(u+\tfrac{1}{u}\right)\left(1+u\,F(u)+\tfrac{F(u)-f}{u}\right)$$
$$+2\left(F(u)-\tfrac{1}{u}\,\tfrac{F(u)-f}{u}\right)\left(1+u\,F(u)+\tfrac{F(u)-f}{u}\right)$$

cancels

Let $U \equiv U(t)$ be the unique fps s.t. $U = 2t\,(U^2+1)\left(1+U\,F(U)+\tfrac{F(U)-f}{U}\right)$

U exists and has positive integer coeffs

then the series U, $V = F(U)$ and $W = \frac{F(U)-f}{U}$ satisfy the system:

$$\begin{cases} U = 2t\,(U^2+1)\,(1+U\,V+W) \\ 0 = U\,V - W \\ V = t\,(1+U\,V+W)^2 \end{cases} \Rightarrow \begin{cases} U = 2t\,(U^2+1)\,(1+2\,U\,V) \\ V = t\,(1+2\,U\,V)^2 \\ W = U\,V \end{cases}$$

and $f = V - U\cdot W$, by definition of U, V, W.

Equations for Dyck trees are particularly simple!

The system can be further simplified

$$\begin{cases} U &= 2t\left(U^2+1\right)(1+2\,U\,V) \\ V &= t\,(1+2\,U\,V)^2 \end{cases}$$

Equations for Dyck trees are particularly simple!

The system can be further simplified

$$\begin{cases} U &= 2t\left(U^2+1\right)(1+2\,U\,V) \\ V &= t\,(1+2\,U\,V)^2 \end{cases} \Rightarrow \begin{cases} U &= 2tU\,(1+2\,U\,V)\cdot U + 2t\,(1+2\,U\,V) \\ V &= 2tU\,(1+2\,U\,V)\cdot V + t\,(1+2\,U\,V) \end{cases}$$

$$\Rightarrow \begin{cases} U &= \frac{2t\,(1+2U\,V)}{1-2tU\,(1+2U\,V)} \\ V &= \frac{t\,(1+2U\,V)}{1-2tU\,(1+2U\,V)} \end{cases} \Rightarrow \quad U = 2\,V$$

Equations for Dyck trees are particularly simple!

The system can be further simplified

$$\begin{cases} U = 2t\left(U^2+1\right)(1+2\,U\,V) \\ V = t\,(1+2\,U\,V)^2 \end{cases} \Rightarrow \begin{cases} U = 2tU\,(1+2\,U\,V)\cdot U + 2t\,(1+2\,U\,V) \\ V = 2tU\,(1+2\,U\,V)\cdot V + t\,(1+2\,U\,V) \end{cases}$$

$$\Rightarrow \begin{cases} U = \frac{2t\,(1+2U\,V)}{1-2tU\,(1+2U\,V)} \\ V = \frac{t\,(1+2U\,V)}{1-2tU\,(1+2U\,V)} \end{cases} \Rightarrow \quad U = 2\,V$$

Theorem: The gf $f = f(t)$ of Dyck trees with n vertices satisfies:

$$f = V - 4V^3 \qquad \text{where} \qquad V = t(1+4V^2)^2$$

Equations for Dyck trees are particularly simple!

Theorem: The gf $f = f(t)$ of Dyck trees with n vertices satisfies:

$$f = V - 4V^3 \qquad \text{where} \qquad V = t(1 + 4V^2)^2$$

Equations for Dyck trees are particularly simple!

Theorem: The gf $f = f(t)$ of Dyck trees with n vertices satisfies:

$$f = V - 4V^3 \qquad \text{where} \qquad V = t(1 + 4V^2)^2$$

using Lagrange inversion theorem:

$$[t^n]V = \frac{1}{n}[x^{n-1}](1 + 4x^2)^{2n} = \begin{cases} \frac{4^m}{2m+1}\binom{4m+2}{m} & \text{if } n = 2m + 1, \\ 0 & \text{otherwise} \end{cases}$$

Equations for Dyck trees are particularly simple!

Theorem: The gf $f = f(t)$ of Dyck trees with n vertices satisfies:

$$f = V - 4V^3 \qquad \text{where} \qquad V = t(1+4V^2)^2$$

using Lagrange inversion theorem:

$$[t^n]V = \frac{1}{n}[x^{n-1}](1+4x^2)^{2n} = \begin{cases} \frac{4^m}{2m+1}\binom{4m+2}{m} & \text{if } n = 2m+1, \\ 0 & \text{otherwise} \end{cases}$$

and

$$[t^n]f = \frac{1}{n}[x^{n-1}](x-4x^3)'(1+4x^2)^{2n} = \begin{cases} \frac{4^m}{(m+1)(2m+1)}\binom{4m+2}{m} & \text{if } n = 2m+1, \\ 0 & \text{otherwise} \end{cases}$$

Marking and identification of V

Theorem: The gf $f = f(t)$ of Dyck trees with n vertices satisfies:

$$f = V - 4V^3 \qquad \text{where} \qquad V = t(1+4V^2)^2$$

using Lagrange inversion theorem:

$$[t^n]V = \frac{1}{n}[x^{n-1}](1+4x^2)^{2n} = \begin{cases} \frac{4^m}{2m+1}\binom{4m+2}{m} & \text{if } n = 2m+1, \\ 0 & \text{otherwise} \end{cases}$$

and

$$[t^n]f = \frac{1}{n}[x^{n-1}](x-4x^3)'(1+4x^2)^{2n} = \begin{cases} \frac{4^m}{(m+1)(2m+1)}\binom{4m+2}{m} & \text{if } n = 2m+1, \\ 0 & \text{otherwise} \end{cases}$$

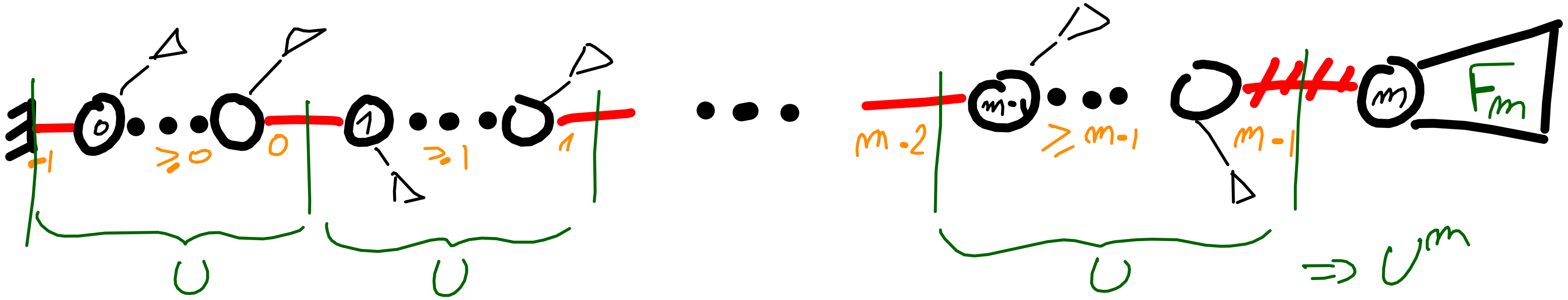
Observe that $\qquad [t^{2m+1}]V = (m+1)[t^{2m+1}]f = [t^{2m+1}]f^{-\bullet}$

$\Rightarrow V$ is the gf of (rooted) Dyck trees with a marked red edge

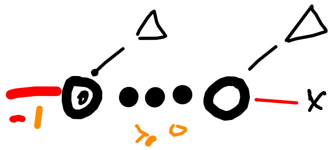
Last passage decomposition and identification of U

The series V is the gf of (rooted) Dyck trees with a marked red edge

Consider a Łukasiewicz (or last passage) factorization of the weight sequence along the branch toward the root.



Now recall we defined $V = F(U) = \sum_{m \geq 0} U^m [u^m] F(u)$

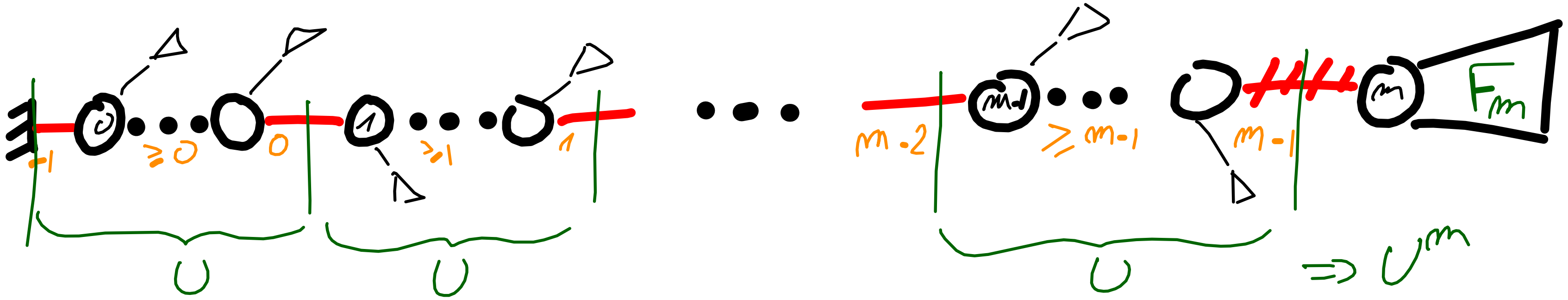
so that $U =$ 

$\Rightarrow$ our series U is the gf of Dyck trees with a marked leaf !

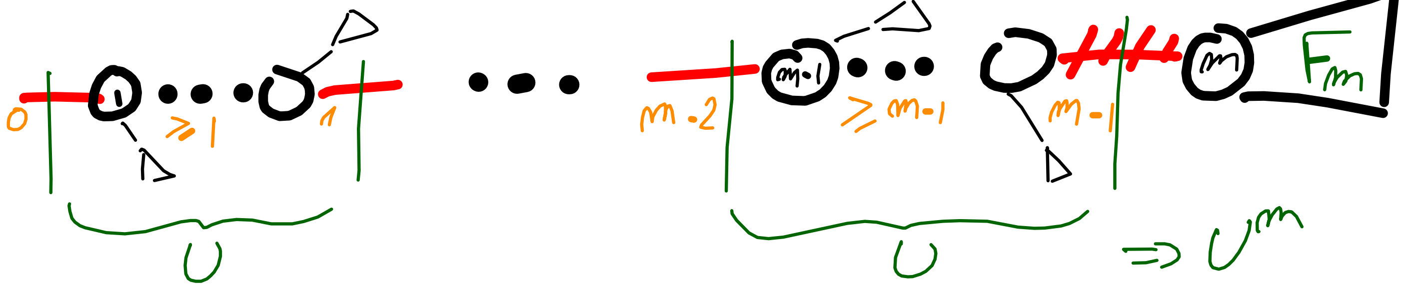
The core of a balanced tree and identification of W

The series V is the gf of (rooted) Dyck trees with a marked red edge
The series U is the gf of Dyck trees with a marked leaf

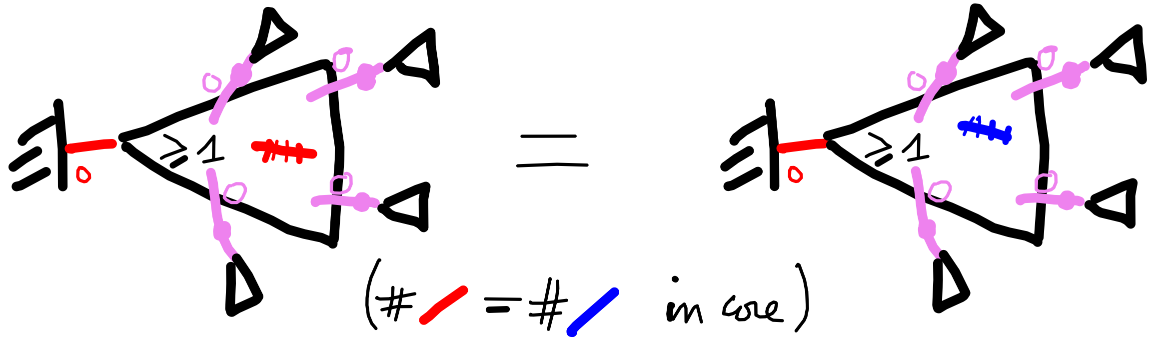
Recall $V = \sum_{m \geq 0} U^m F_m =$



hence $W = \sum_{m \geq 1} U^{m-1} F_m =$



$=$

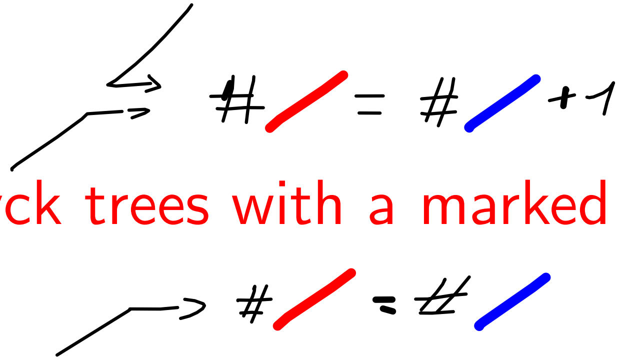


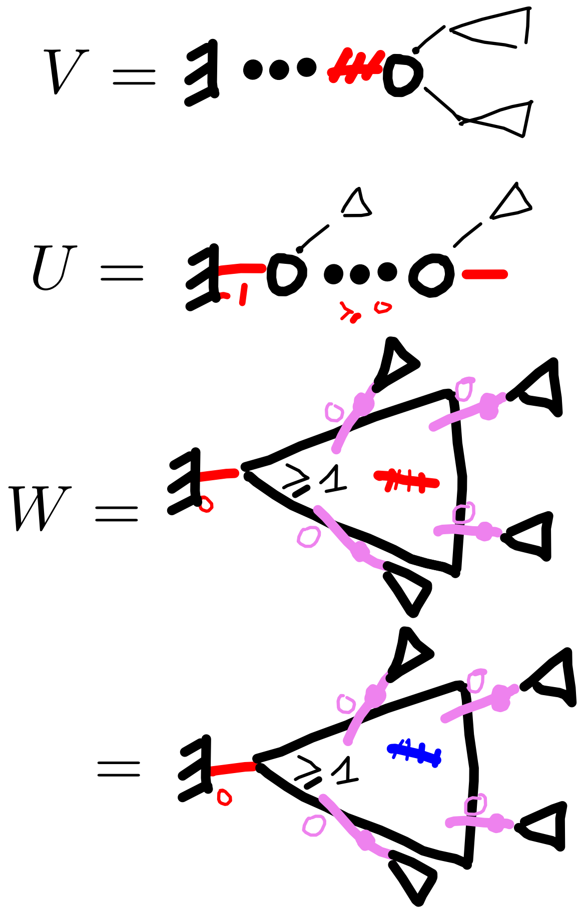
$\Rightarrow$ W is the gf of balanced positive trees with a marked blue edge in their internally positive core.

Decomposing marked Dyck-Łukasiewicz trees

Let's now restart from the combinatorial interpretations: let

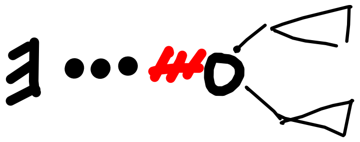
- V denote the gf of (rooted) Dyck trees with a marked red edge
- U denote the gf of Dyck trees with a marked leaf
- W denote the gf of balanced positive trees with a marked red edge in their internally positive core.
  W is also the gf of balanced positive trees with a marked blue edge in their internally positive core.



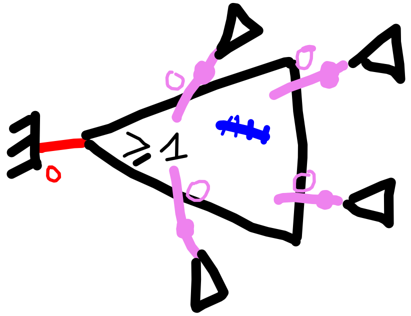


We would like a direct quaternary decomposition of these marked rooted trees to reprove directly that $V = t(1 + 4V^2)^2$.

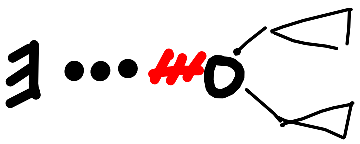
Combinatorial derivation of $U = 2V$ and $W = U\,V$

$V =$ 

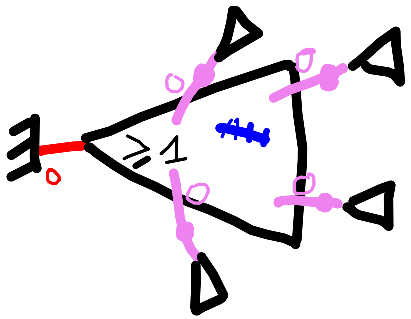
$U =$ 

$W =$ 

Combinatorial derivation of $U = 2V$ and $W = U\,V$

$V =$ 

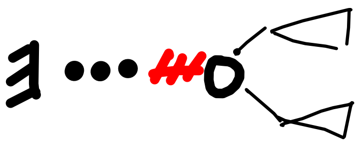
$U =$ 

$W =$ 

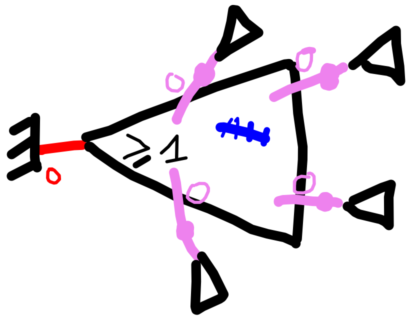
Claim: There is a 2-to-1 correspondance between Dyck trees with a marked leaf and Dyck trees with a marked red edge with the same size

Immediate since a Dyck tree with $2n+1$ vertices has $n+1$ red edges and $2n+2$ leaves.

Combinatorial derivation of $U = 2V$ and $W = U\,V$

$V =$ 

$U =$ 

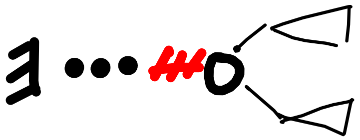
$W =$ 

Claim: There is a 2-to-1 correspondance between Dyck trees with a marked leaf and Dyck trees with a marked red edge with the same size

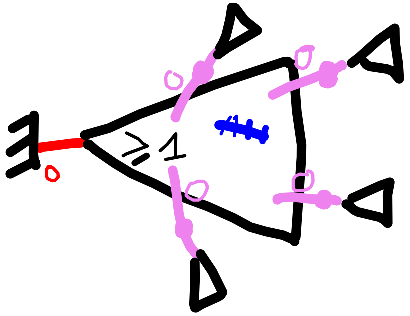
$\Rightarrow U = 2V$

Immediate since a Dyck tree with $2n+1$ vertices has $n+1$ red edges and $2n+2$ leaves.

Combinatorial derivation of $U = 2V$ and $W = U\,V$

$V =$ 

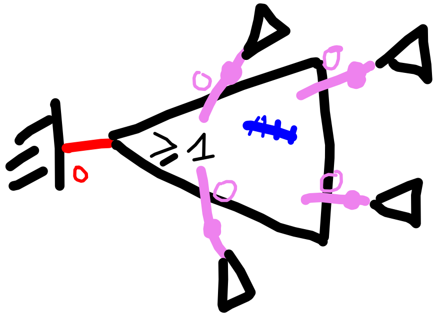
$U =$ 

$W =$ 

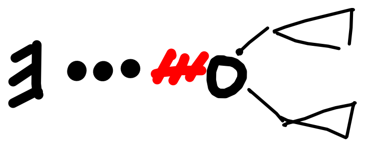
Claim: There is a 2-to-1 correspondance between Dyck trees with a marked leaf and Dyck trees with a marked red edge with the same size

$\Rightarrow U = 2V$

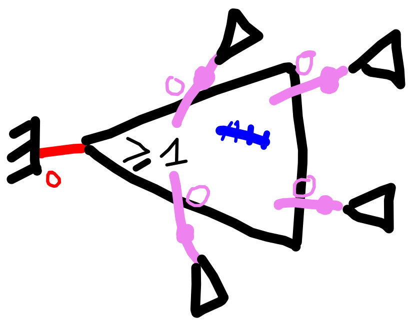
Immediate since a Dyck tree with $2n+1$ vertices has $n+1$ red edges and $2n+2$ leaves.

$W =$ 

Combinatorial derivation of $U = 2V$ and $W = U\,V$

$V =$ 

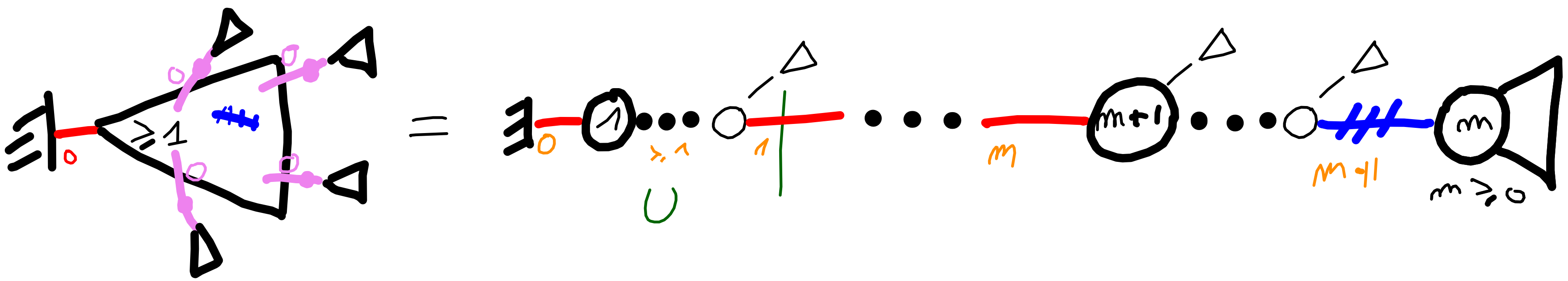
$U =$ 

$W =$ 

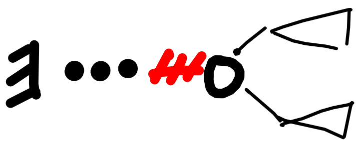
Claim: There is a 2-to-1 correspondance between Dyck trees with a marked leaf and Dyck trees with a marked red edge with the same size

$\Rightarrow U = 2V$

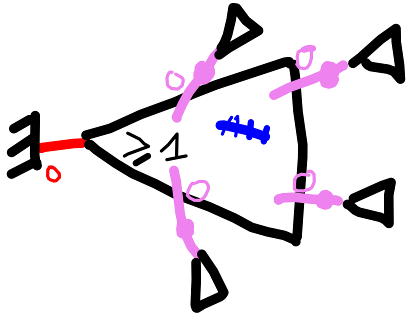
Immediate since a Dyck tree with $2n+1$ vertices has $n+1$ red edges and $2n+2$ leaves.

$W =$ 

Combinatorial derivation of $U = 2V$ and $W = U\,V$

$V =$ 

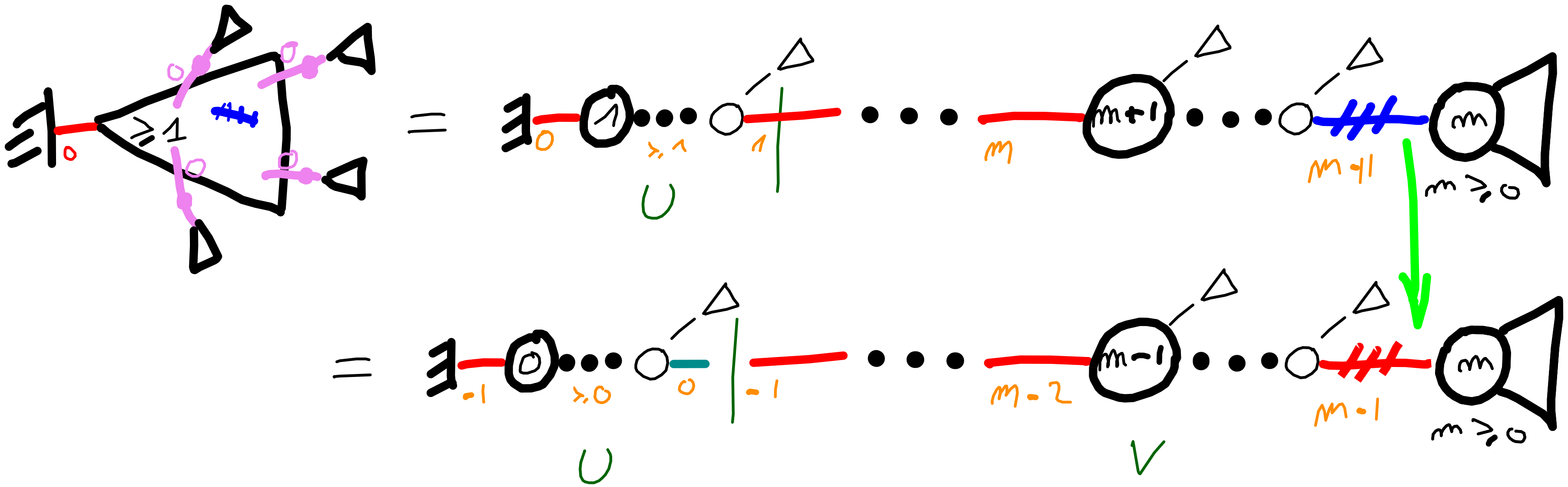
$U =$ 

$W =$ 

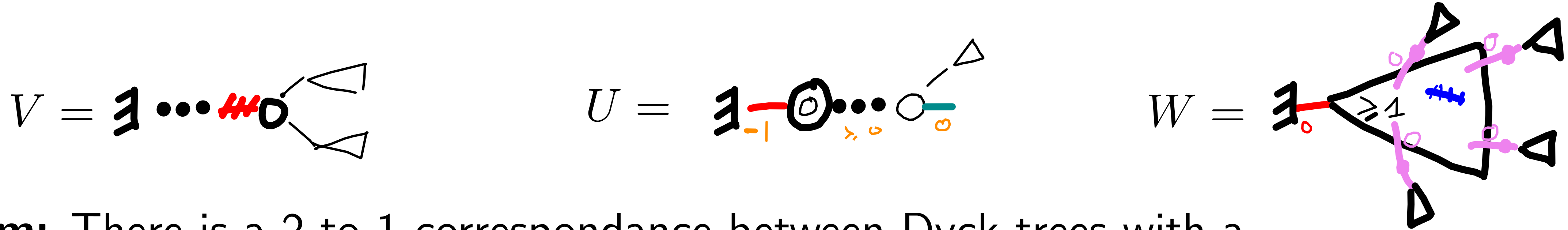
Claim: There is a 2-to-1 correspondance between Dyck trees with a marked leaf and Dyck trees with a marked red edge with the same size

$\Rightarrow U = 2V$

Immediate since a Dyck tree with $2n+1$ vertices has $n+1$ red edges and $2n+2$ leaves.

$W =$ 

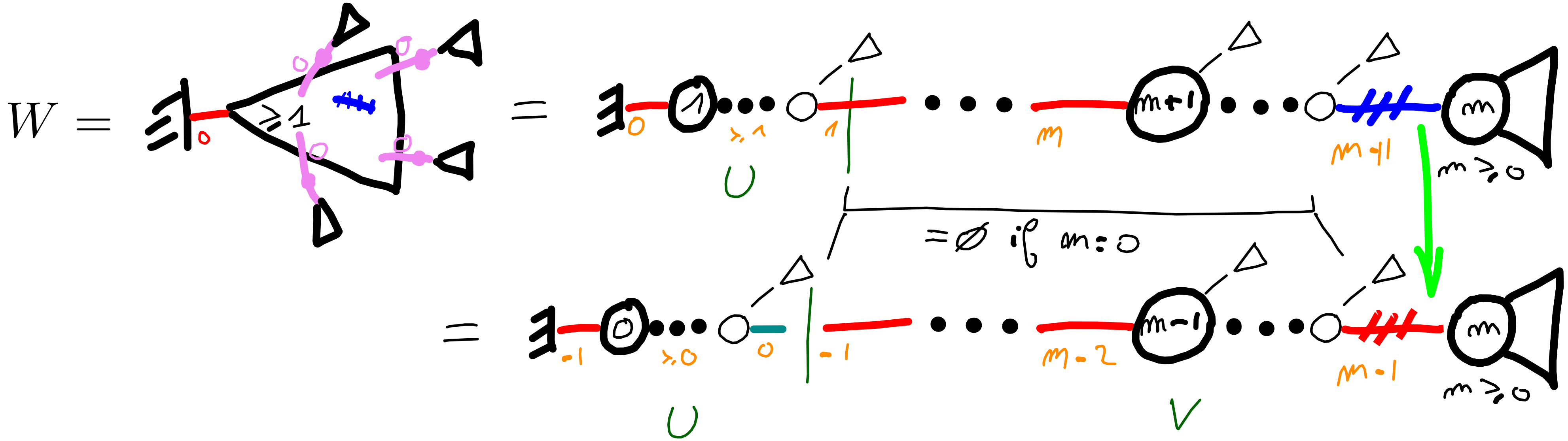
Combinatorial derivation of $U = 2V$ and $W = U\,V$



Claim: There is a 2-to-1 correspondance between Dyck trees with a marked leaf and Dyck trees with a marked red edge with the same size

$\Rightarrow U = 2V$

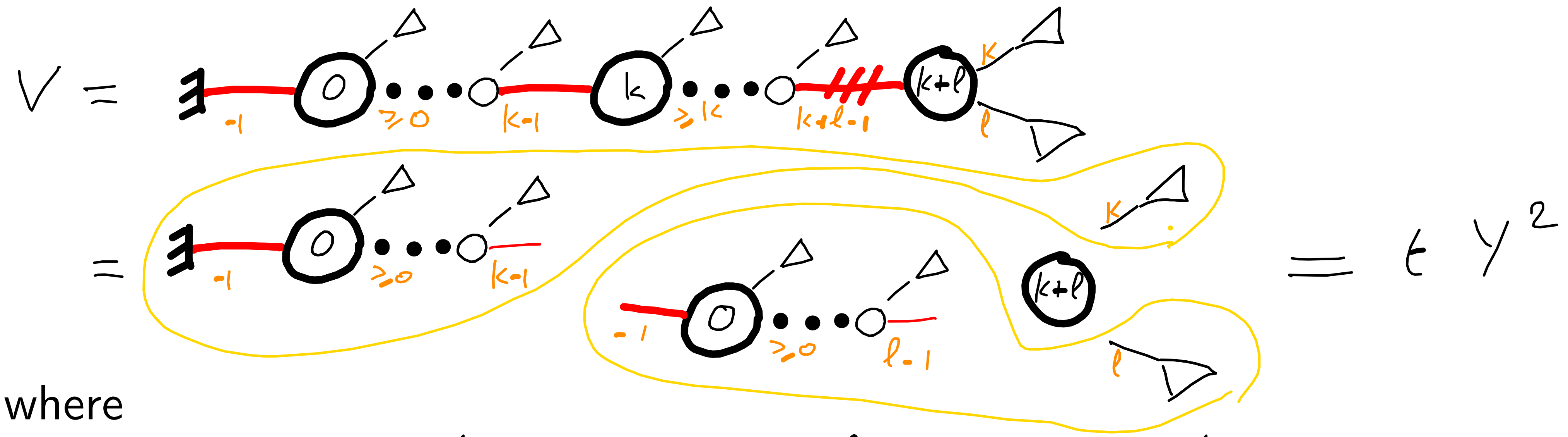
Immediate since a Dyck tree with $2n+1$ vertices has $n+1$ red edges and $2n+2$ leaves.



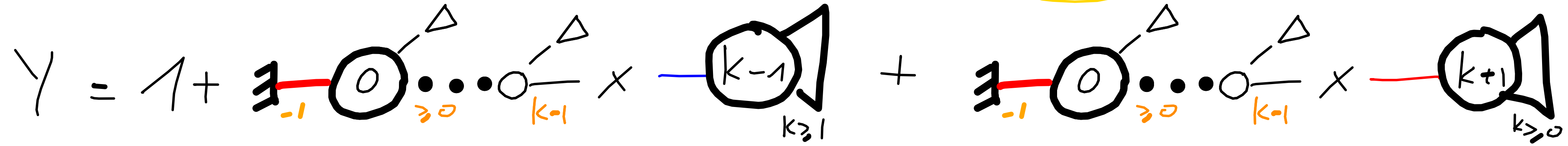
$\Rightarrow W = U\,V$

Finally, a quaternary decomposition of marked Dyck trees

Theorem: The class of marked Dyck trees admit the following decomposition:

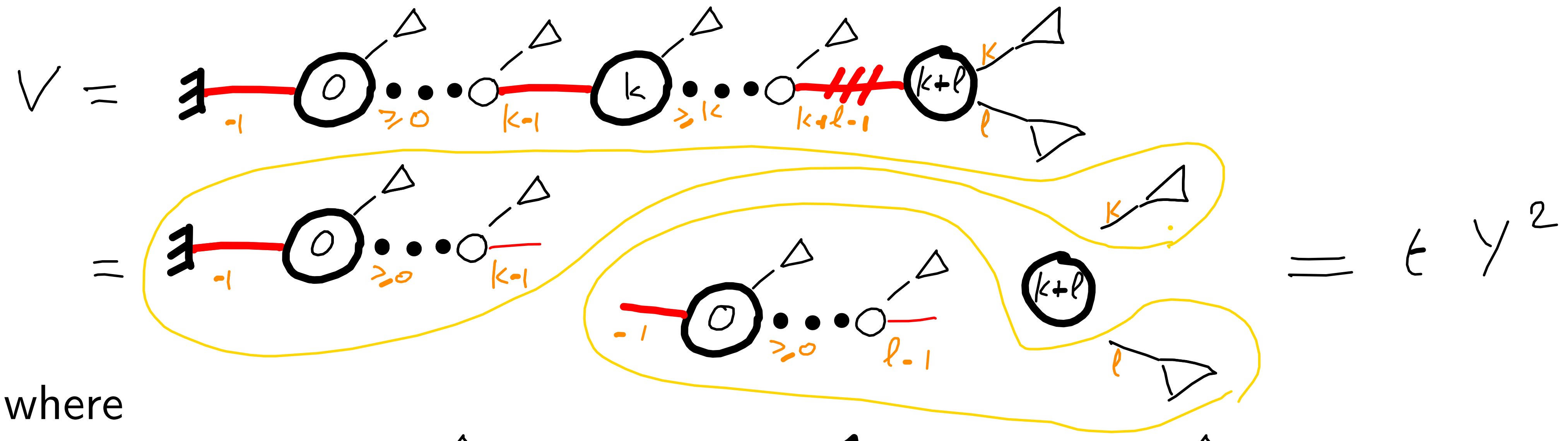


where

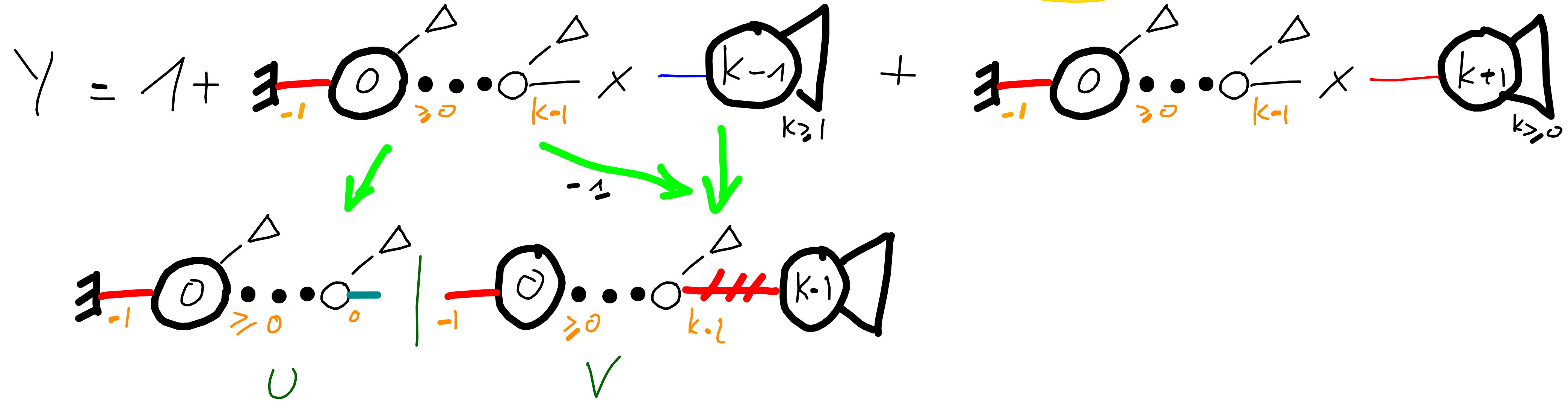


Finally, a quaternary decomposition of marked Dyck trees

Theorem: The class of marked Dyck trees admit the following decomposition:

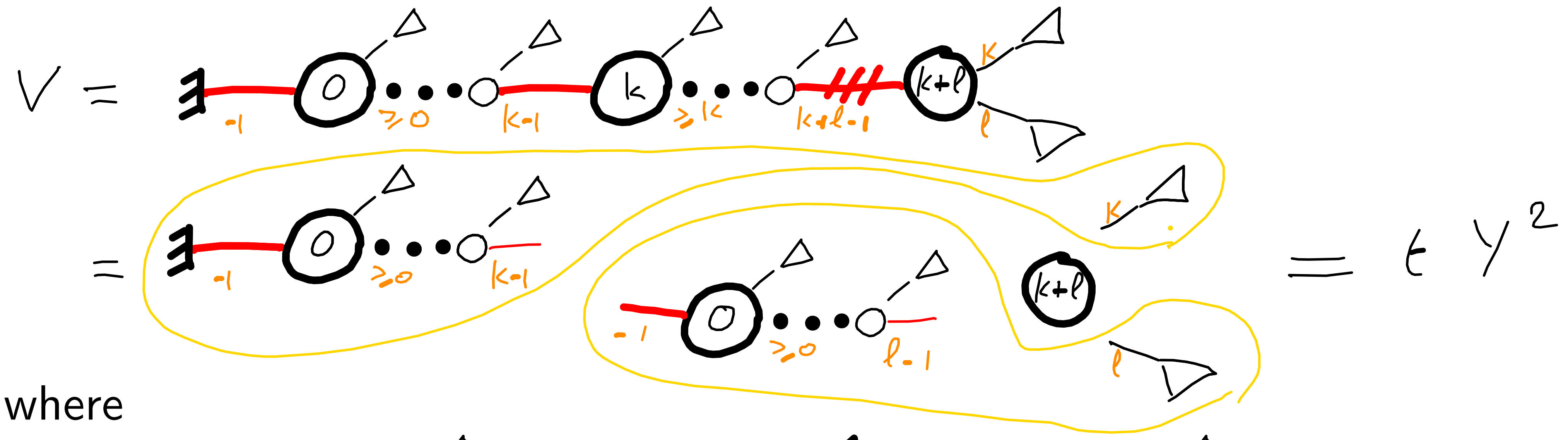


where

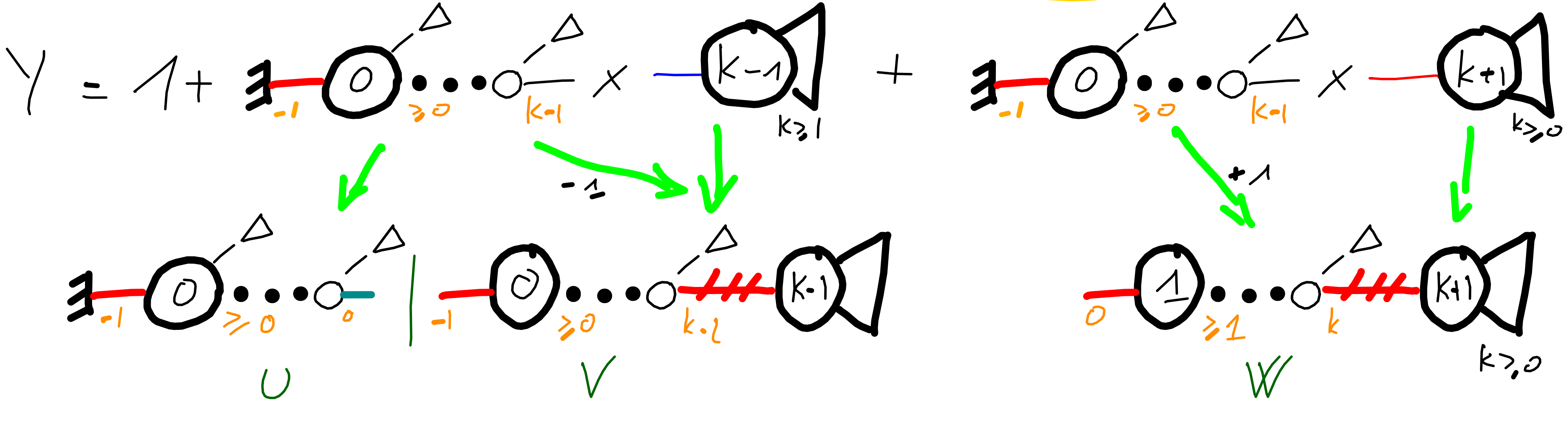


Finally, a quaternary decomposition of marked Dyck trees

Theorem: The class of marked Dyck trees admit the following decomposition:

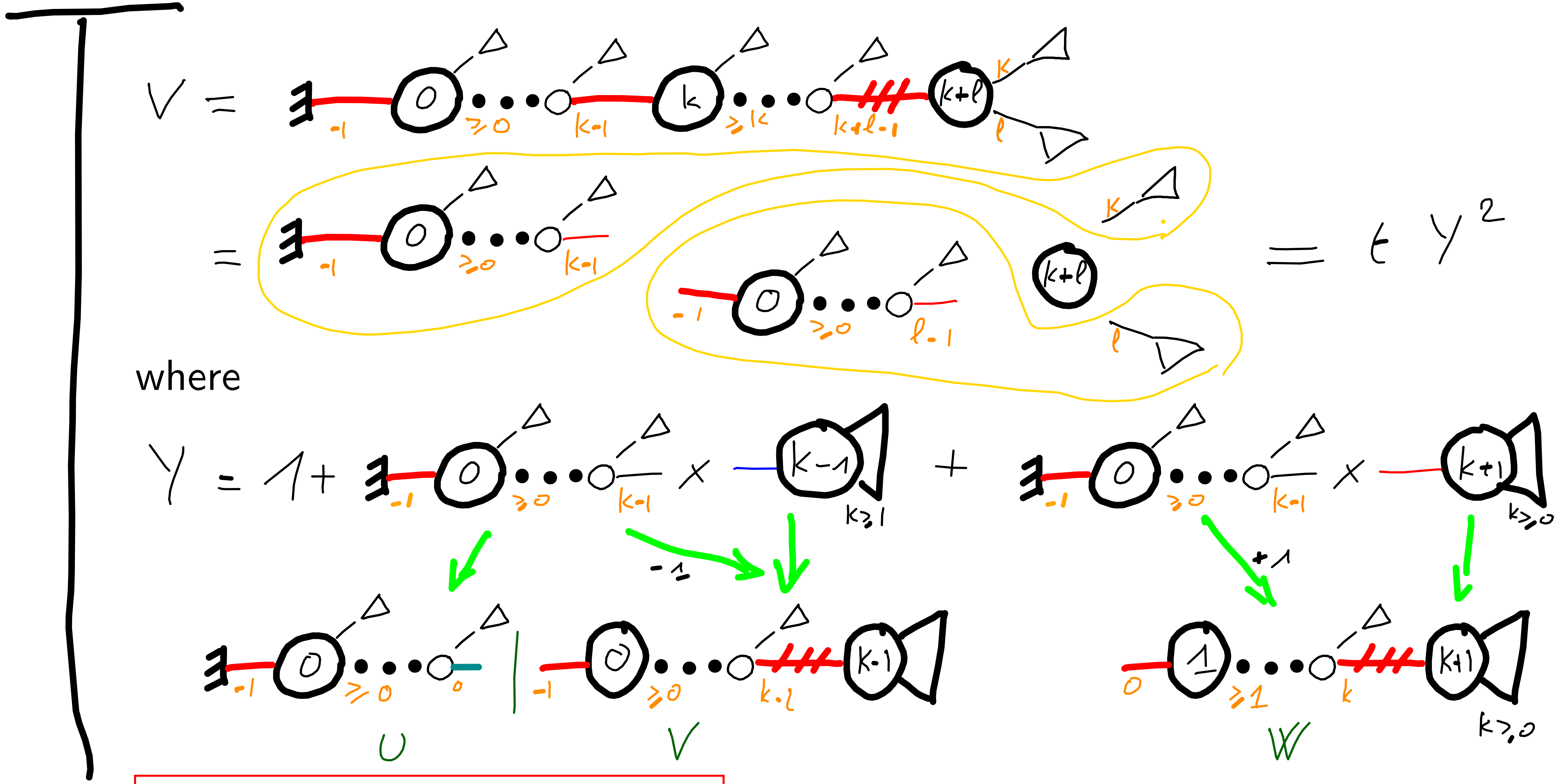


where



Finally, a quaternary decomposition of marked Dyck trees

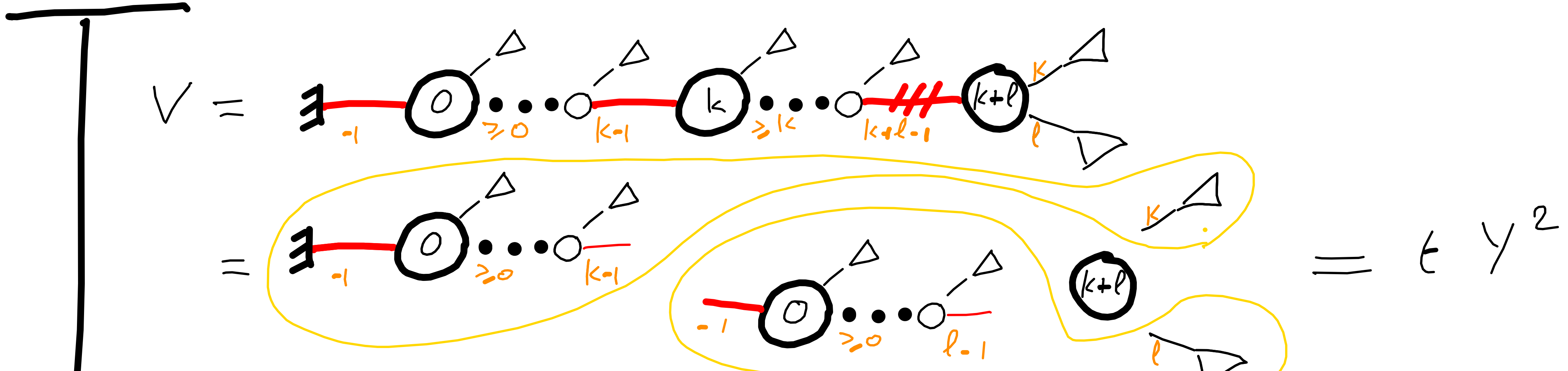
Theorem: The class of marked Dyck trees admit the following decomposition:



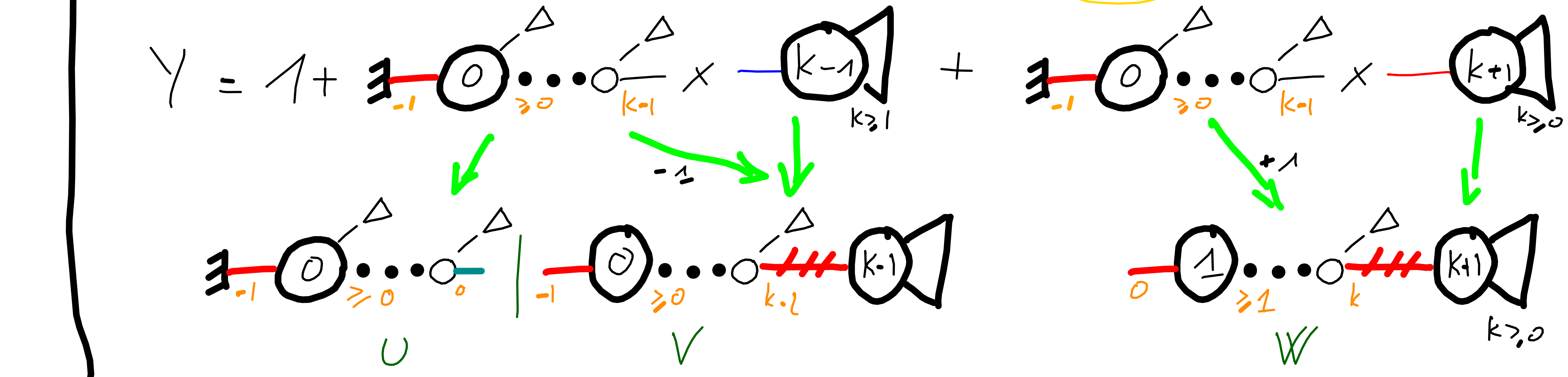
$$\Rightarrow V = t\,(1 + U\,V + W)^2$$

Finally, a quaternary decomposition of marked Dyck trees

Theorem: The class of marked Dyck trees admit the following decomposition:



where



$\Rightarrow V = t\,(1 + U\,V + W)^2$

$\Rightarrow V = t\,(1 + 2 \cdot 2V \cdot V)^2$

Complexity of random generation and a resampling trick

Random generation

Theorem[Sportiello 21]: Linear time random generation for context free classes.

$$\Rightarrow V = t\,(1 + 2 \cdot 2\,V \cdot V)^2$$

A class of rooted (2+2)-ary trees...

Sportiello's theorem allows to generate the decomposition trees in linear time.

However the intermediate transformations on Dyck-Łukasiewicz trees have a priori an extra linear cost:

$$V = t\,(1 + U\,V + W)^2$$ with $U = 2V$ and $W = U \cdot V$.

leaf $\leftrightarrow$ marked red

marked blue $\leftrightarrow$ marked red

Random generation and complexity

Theorem[Resampling trick]: from a uniform random $(2+2)$-ary tree, reconstruct in (quasi-)linear time a resampled Dyck-Łukasiewicz tree

Sportiello (or direct encoding) yields: $$V = t\,(1 + 2 \cdot 2\,V \cdot V)^2$$

$\Rightarrow \quad V = t\,(1 + U\,V + W)^2$ with $U = 2V$ and $W = U \cdot V$.

select random leaf (for $U = 2V$)

select random blue edge (for $W = U \cdot V$)

all other grafting operations can be done in (quasi-)constant time (constant number of pointer operation and small integer sampling).

Conclusion

Extend to polynomial equations with one catalytic variable

Let $Q(v, w, u) = \sum_{i,j,k \geq 0} q_{ijk} v^i w^j u^k$ a formal power series

Let $F(u) \equiv F(u, a, b, t)$ the unique fps* solution of

$$F(u) = t\,Q\left(F(u), \frac{b}{u}(F(u) - f), a\,u\right) \qquad \text{with } f = F(0)$$

Then the derivative f'_t satisfies a system of positive algebraic equation.

$$\begin{cases} V &= t\,Q(V, bW, aU) \\ U &= t\,U\,Q'_v(V, bW, aU) + t\,b\,Q'_w(V, bW, aU) \\ W &= t\,W\,Q'_v(V, bW, aU) + t\,aQ'_u(V, bW, aU) \\ (tf'_t) &= t\,(tf'_t)\,Q'_v(V, bW, aU) + V \end{cases}$$

with a full combinatorial interpretation that allows for random generation using Sportiello's general approach for context free structures and resampling.

Application of the general result

Special cases: this yields algebraic decompositions for

- Linxiao Chen's fully parked trees (2021)
- Duchi et al.'s fighting fish and variants (2016)
- Various families of permutations (West's two-stack sortable) (1990)
- Tutte's map decomposition (60's)

Works as well with exponential series: Dyck Cayley trees.

However in most of the cases combinatorial intuition is still needed to simplify the resulting decompositions, and express it in terms of the original structures.

Thank you!