# Image interdisciplinaire de la Combinatoire Analytique

## Interdisciplinary image of Analytic Combinatorics

**Sergey Dovgal**

Laboratoire d'Informatique de Paris Nord,
Université Paris-13

Sous direction de
**Olivier Bodini** et **Vlady Ravelomanana**

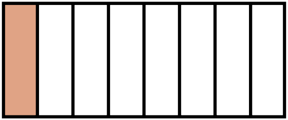
# Why analytic combinatorics

# Outline of the current talk

$$\underbrace{\text{Interdisciplinary image of } \underbrace{\text{Analytic Combinatorics}}_{\text{Part I}}}_{\text{Part II}}$$

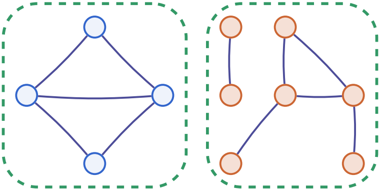
# Part I. Analytic Combinatorics

## Level I. Symbolic method



*(You can never be too thorough)*

# Known convolution rules



| convolution | generating function | name |
|---|---|---|
| $\sum_{k=0}^{n} a_k b_{n-k}$ | $A(z) := \sum_{n\geq 0} a_n z^n$ | ordinary GF |
| $\sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}$ | $A(z) := \sum_{n\geq 0} a_n \frac{z^n}{n!}$ | exponential GF |
| $\sum_{d\mid n} a_d b_{n/d}$ | $A(s) := \sum_{n\geq 1} \frac{a_n}{n^s}$ | Dirichlet series |
| $\sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k} 2^{k(n-k)}$ | $A(z) := \sum_{n\geq 0} a_n \frac{z^n}{2^{\binom{n}{2}} n!}$ | graphic GF |

# Combinatorial operations

| Operation | interpretation |
|---|---|
| $A(z) + B(z)$ | disjoint union |
| $A(z)B(z)$ | cartesian product |
| $\exp(A(z))$ | set |
| $A(B(z))$ | substitution |
| $z\partial_z A(z)$ | pointing |
| $A(z) \odot B(z)$ | Hadamard product |
| $\dots$ | $\dots$ |

**The philosophy of the symbolic method.** (Bergeron, Labelle, Leroux).

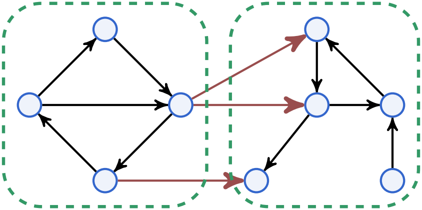
Decomposition $\Rightarrow$ Equation

**Follow-up: asymptotic analysis.** (Flajolet, Odlyzko, ...).

Equation $\Rightarrow$ GF expansion $\Rightarrow$ Asymptotics

# Graphic convolution rule

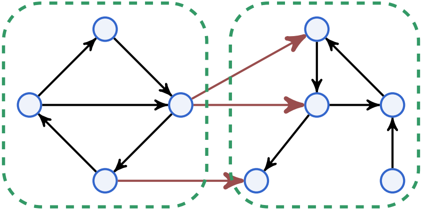
$$A(z) = a_0 + \frac{a_1 z}{1!2^{\binom{1}{2}}} + \frac{a_2 z^2}{2!2^{\binom{2}{2}}} + \ldots, \quad B(z) = b_0 + \frac{b_1 z}{1!2^{\binom{1}{2}}} + \frac{b_2 z^2}{2!2^{\binom{2}{2}}} + \ldots$$



| **coefficient level** | **graphic GF level** |
|---|---|
| $c_n := \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k} 2^{k(n-k)}$ | $C(z) = A(z) \cdot B(z)$ |

# Graphic convolution rule

$$A(z, w) = a_0 + \frac{a_1(w) z}{1!(1+w)^{\binom{1}{2}}} + \frac{a_2(w) z^2}{2!(1+w)^{\binom{2}{2}}} + \frac{a_3(w) z^3}{3!(1+w)^{\binom{3}{2}}} + \dots$$



| coefficient level | graphic GF level |
|---|---|
| $c_n := \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k} 2^{k(n-k)}$ | $C(z) = A(z) \cdot B(z)$ |
| $c_n := \sum_{k=0}^{n} \binom{n}{k} a_k(w) b_{n-k}(w) (1+w)^{k(n-k)}$ | $C(z, w) = A(z, w) B(z, w)$ |

$w$ marks edges

# Enumerative result

[De Panafieu, D. '19]

*Exponential Hadamard product:*

$$\left(\sum_{n\geq 0} a_n \frac{z^n}{n!}\right) \odot \left(\sum_{n\geq 0} b_n \frac{z^n}{n!}\right) := \sum_{n\geq 0} a_n b_n \frac{z^n}{n!}$$

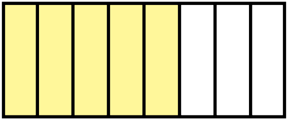
*Exponential GF for graphs, graphic GF for sets:*

$$\mathbb{G}(z,w) = \sum_{n\geq 0} (1+w)^{\binom{n}{2}} \frac{z^n}{n!}, \quad \widehat{Set}(z,w) = \sum_{n\geq 0} \frac{1}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!},$$

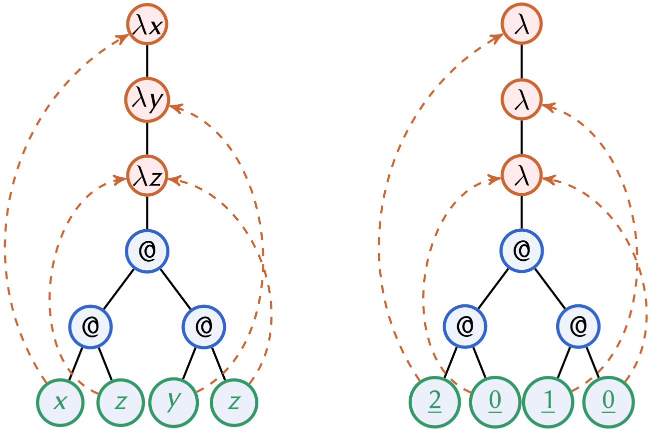
**Theorem.** (Rediscovery of results by Liskovets and Robinson)

| | **digraphs** | **graphs** |
|---|---|---|
| EGF, connected | $-\log\left(\mathbb{G}(z,w) \odot \frac{1}{\mathbb{G}(z,w)}\right)$ | $-\log\left(\frac{1}{\mathbb{G}(z,w)}\right)$ |
| components from given family $\mathcal{C}$ | $GGF = \frac{1}{\widehat{Set}(z,w) \odot e^{-C(z,w)}}$ | $EGF = \frac{1}{e^{-C(z,w)}}$ |

Level II. Systems of algebraic equations



# Lambda terms




A tree-like representation of a *closed* lambda term
$\lambda x.\lambda y.\lambda z.(xz)(yz)$

# Drmota–Lalley–Woods theorem (simplified)

**Theorem.**

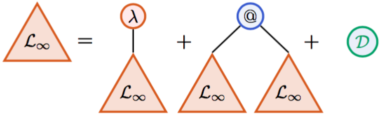
- Let $F(z)$ be a generating function satisfying

$$F(z) = \Phi(F(z), z)$$

  with $\Phi$ is nonlinear, strongly-connected
  and has "combinatorial origin"
- Then, there exists $\rho > 0$ such that as $z \to \rho$, $z \in \mathbb{C}$

$$F(z) \sim a - b\sqrt{1 - \frac{z}{\rho}}$$

**Example.** Plain lambda terms:




$$L(z) = zL(z) + zL^2(z) + \frac{z}{1-z}$$

# Drmota–Lalley–Woods theorem in infinite dimension

$$F(z, \vec{u}) = \Phi(F(z, \vec{u}), z, \vec{u})$$

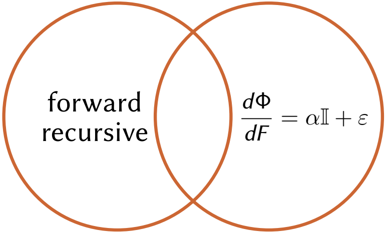
**Theorem.** [Drmota, Gittenberger, Morgenbesser '2012] simplified

Let $\dim \Phi = \infty$. If the Jacobian $\dfrac{d\Phi}{dF}$ can be represented as $\alpha \cdot \mathbb{I} + \varepsilon$, where $\varepsilon$ is a compact operator, then DLW expansion holds:

$$F(z) \sim a(\vec{u}) - b(\vec{u})\sqrt{1 - \frac{z}{\rho(\vec{u})}}$$

**Theorem.** [Bendkowski, Bodini, D. '2019] simplified

If $\Phi$ is $\infty$-dimensional *forward-recursive* then DLW expansion holds.



# Formulation of our result

[Bendkowski, Bodini, D.]

**Definition.** (Forward recursive systems)

$$\vec{L}_m = \vec{K}_m(\vec{L}_m, \vec{L}_{m+1}, z, \vec{u}), \quad m = 0, 1, 2, \ldots, \infty.$$

satisfies the conditions

- $\vec{K}_m(\vec{\ell}_1, \vec{\ell}_2, z, \vec{u}) \succeq \vec{K}_\infty(\vec{\ell}_1, \vec{\ell}_2, z, \vec{u})$
- $\vec{K}_\infty$ is DLW-well-defined
- $\vec{K}_\infty(\vec{L}_\infty, \vec{L}_\infty, z, \vec{u}) - \vec{K}_m(\vec{L}_\infty, \vec{L}_\infty, z, \vec{u}) \preceq \vec{A}(z, \vec{u}) \cdot \underbrace{B(z, \vec{u})^m}_{<1 \text{ at } z=\rho,\ \vec{u}=\vec{1}}$ ;

**Theorem.** (Coefficient transfer)

$$[z^n]\vec{L}_m(z, \vec{u}) \sim [z^n] \sum_{k \geq 0} \vec{c}_{m,k} \left(1 - \frac{z}{\rho(\vec{u})}\right)^{k/2}$$

$$\|\vec{c}_{m,k} - \vec{c}_{\infty,k}\| \sim C^{-m}$$

# Illustration of a forward-recursive system

**Example.** *Closed* lambda terms.

$$
\begin{aligned}
L_0(z) &= zL_1(z) + zL_0(z)^2 \quad , \\
L_1(z) &= zL_2(z) + zL_1(z)^2 + z \quad , \\
L_2(z) &= zL_3(z) + zL_2(z)^2 + z + z^2 \quad , \\
&\dots \\
L_\infty(z) &= zL_\infty(z) + zL_\infty(z)^2 + \frac{z}{1-z} \quad .
\end{aligned}
$$

**Conditions to check.**

- $z\lambda_1 + z\lambda_2^2 + z + z^2 + \ldots + z^m \preceq z\lambda_1 + z\lambda_2^2 + \dfrac{z}{1-z}$
- $z^{m+1} + \ldots \preceq \alpha(z) \cdot \beta(z)^m$
- $L_\infty(z)$ satisfies DLW conditions

**Generalisation.** Decorate with marking variables.

# Multiparametric sampling

Context-free grammar with terminals $\bullet_1, \bullet_2, \bullet_3, \bullet_4$ and non-terminals $S_1, \ldots, S_n$:

$$S_i \to \sum_j T_{ij}(S_1, \ldots, S_n, \bullet_1, \bullet_2, \bullet_3, \bullet_4)$$

**Exact sampling.** Given positive integers $(n_1, n_2, \ldots, n_d)$, sample *uniformly at random* a word $w \in S_1$ such that

$$|w|_{\bullet_1} = n_1, \quad |w|_{\bullet_2} = n_2, \quad \cdots, \quad |w|_{\bullet_d} = n_d$$

:( This problem is #P-complete ):

**Boltzmann sampling.** Given positive reals $(z_1, z_2, \ldots, z_d)$, sample a word $w \in S_1$ so that

$$\mathbb{P}(w\colon |w|_{\bullet_1} = n_1, |w|_{\bullet_2} = n_2, \cdots, |w|_{\bullet_d} = n_d) = \frac{z_1^{n_1} z_2^{n_2} \cdots z_d^{n_d}}{\sum_{n_1,\ldots,n_d} z_1^{n_1} z_2^{n_2} \cdots z_d^{n_d}}$$

(: There exists a polynomial algorithm :)

# Boltzmann sampler

$$S_i \to \sum_j T_{ij}(S_1, \ldots, S_n, \bullet)$$

**Algorithm 1:** Boltzmann sampler for context-free grammars

**Data:** real value $z > 0$

**Result:** Random word from Boltzmann distribution

**Function** $\Gamma R_i(z)$:

  **if** *$R_i$ is terminal* **then**

    **return** $\bullet$ ;

  **for** *all $j$* **do**

    $p_j := \dfrac{T_{ij}(S_1(z), \ldots, S_n(z), z)}{R_i(z)}$ ;

  Choose the transition $T_{ij}$ with probability $p_j$ ;

  $A_1 A_2 \ldots A_k := T_{ij}$ ;

  **return** $\Gamma A_1(z) \Gamma A_2(z) \cdots \Gamma A_k(z)$ ;

# Multiparametric Boltzmann sampler

$$S_i \to \sum_j T_{ij}(S_1, \ldots, S_n, \bullet_1, \bullet_2, \cdots, \bullet_\ell)$$

**Algorithm 2:** Boltzmann sampler for context-free grammars

**Data:** real values $z_1, z_2, \cdots, z_\ell > 0$

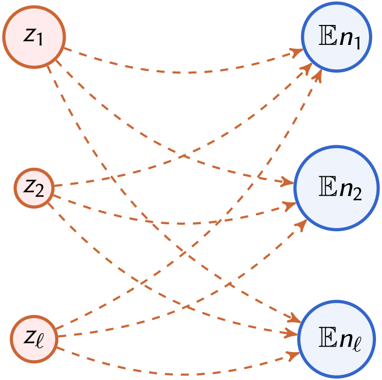
**Result:** Random word from Boltzmann distribution

**Function** $\Gamma R_i(z)$:

- **if** $R_i$ *is terminal* $\bullet_k$ **then**
  - **return** $\bullet_k$ ;
- **for** *all* $j$ **do**
  - $p_j := \dfrac{T_{ij}(S_1(\boldsymbol{z}), \ldots, S_n(\boldsymbol{z}), z_1, z_2, \cdots, z_\ell)}{R_i(\boldsymbol{z})}$ ;
- Choose the transition $T_{ij}$ with probability $p_j$ ;
- $A_1 A_2 \ldots A_k := T_{ij}$ ;
- **return** $\Gamma A_1(\boldsymbol{z}) \Gamma A_2(\boldsymbol{z}) \cdots \Gamma A_k(\boldsymbol{z})$ ;

# Tuning of the multiparametric Boltzmann sampler

Handles $\boldsymbol{z}$ $\Rightarrow$ Expectations $\mathbb{E}n_k$




‼ The handles cannot be tuned independently ‼

**Tuning equation:**

$$z_1 \frac{\partial_{z_1} F(\vec{z})}{F(\vec{z})} = \mathbb{E}n_1, \quad z_2 \frac{\partial_{z_2} F(\vec{z})}{F(\vec{z})} = \mathbb{E}n_2, \quad \ldots, \quad z_\ell \frac{\partial_{z_\ell} F(\vec{z})}{F(\vec{z})} = \mathbb{E}n_\ell$$

# Tuning of the multiparametric Boltzmann sampler

[Bendkowski, Bodini, D.]

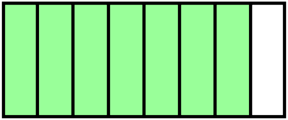
**Theorem.** Let the expected values $n_1, \ldots, n_\ell$ of the terminals $\bullet_1, \bullet_2, \cdots, \bullet_\ell$ be given. Let $S_k(\boldsymbol{z})$ satisfy

$$
\begin{aligned}
S_1 &= \Phi_1(S_1, \ldots, S_n, \boldsymbol{z}),\\
&\ldots\\
S_n &= \Phi_n(S_1, \ldots, S_n, \boldsymbol{z}).
\end{aligned}
$$

The tuning vector $(z_1, \ldots, z_\ell) = (e^{x_1}, \ldots, e^{x_\ell})$ can be obtained by solving a convex optimisation problem

$$
\begin{aligned}
S_1 - n_1 x_1 - n_2 x_2 - \ldots - n_\ell x_\ell &\to \min_{(S_1, \cdots, S_n, x_1, \cdots, x_\ell)},\\
S_1 &\geq \log \Phi_1(e^{S_1}, \ldots, e^{S_n}, e^{x_1}, \ldots, e^{x_\ell}),\\
&\ldots\\
S_n &\geq \log \Phi_n(e^{S_1}, \ldots, e^{S_n}, e^{x_1}, \ldots, e^{x_\ell}).
\end{aligned}
$$

## Level III. Airy function and saddle point



# Boltzmann tuning as saddle point equation

$$[z^n]f(z) = \frac{1}{2\pi i} \oint_{|z|=R} e^{g(z)} \frac{dz}{z}$$

$$\Downarrow$$

$$g'(z) = 0 \quad \Leftrightarrow \quad z\frac{f'(z)}{f(z)} = n$$

$$\Downarrow$$

$$\log f(e^z) - nz \to \min_z \quad \Leftrightarrow \quad \mathbb{E}_z[N] = n$$

$$\left( \mathbb{P}_x(N = k) = \frac{x^k [z^k] f(z)}{f(x)} \right)$$

# Coalescence of saddle points

**Gaussian case.**

$$[z^n]f(z) = \frac{1}{2\pi i}\oint_{|z|=R} e^{g(z)}\frac{dz}{z} \approx \frac{1}{2\pi i}\oint_{|z|=R} e^{g(z_0)+\frac{1}{2}g''(z_0)(z-z_0)}\frac{dz}{z}$$

$$\approx \frac{1}{\sqrt{2\pi g''(z_0)}}e^{g(z_0)}$$

**Coalescence of saddle points.** When $g'(z) = g''(z) = 0$,

$$g(z) \approx g(z_0) + \frac{1}{6}g'''(z_0)(z-z_0)^3$$

**Airy function.**

$$A(y) = \int_{\mathbb{R}} e^{i(x^3/3-xy)}dx, \quad A''(y) + yA(y) = 0$$

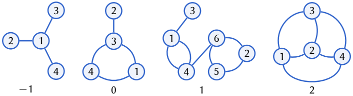
# FJKLP lemma for graphs

[Flajolet, Janson, Knuth, Łuczak, Pittel] → Airy function is hidden inside

**Point of saddle coalescence.** Graphs with $n$ vertices, $m$ edges, $m = \frac{n}{2}(1 - \mu n^{-1/3})$, $\mu \to \infty$:

$$\underset{\substack{\uparrow\\ \text{graphs}}}{\frac{n!}{|\mathcal{G}_{n,m,\Delta}|}} \underset{\substack{\uparrow\\ \text{consist of}}}{\frac{1}{2\pi i}} \oint \underset{\substack{\uparrow\\ \text{trees}}}{\frac{U(z)^{n-m}}{(n-m)!}} \underset{\substack{\uparrow\\ \text{unicycles}}}{e^{V(z)}} \frac{dz}{z^{n+1}} = 1 - \frac{5}{24|\mu|^3} + O(\mu^{-6})$$

**Refined analysis.**



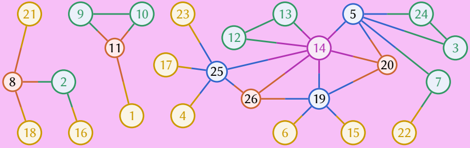


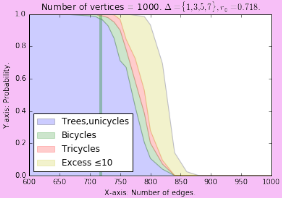
$$\underset{\substack{\uparrow\\ \text{graphs}}}{\frac{n!}{|\mathcal{G}_{n,m,\Delta}|}} \underset{\substack{\uparrow\\ \text{consist of}}}{\frac{1}{2\pi i}} \oint \underset{\substack{\uparrow\\ \text{trees}}}{\frac{U(z)^{n-m}}{(n-m)!}} \underset{\substack{\uparrow\\ \text{unicycles}}}{e^{V(z)}} \underset{\substack{\uparrow\\ \text{complex comp.}}}{\frac{c}{(1-T(z))^{3\gamma}}} \frac{dz}{z^{n+1}} \sim \frac{c}{|\mu|^{3\gamma}}$$

# Graphs with degree constraints

[De Panafieu, Ramos], [D., Ravelomanana]

Allowed degrees: $\Delta = \{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \ldots\}$, $\delta_1 = 1$ or $\delta_2 = 1$.




**EGF of allowed degrees** helps to construct the trees and unicycles, next higher excess

$$\omega(z) = \sum_{d \in \Delta} \frac{z^d}{d!} \quad \Rightarrow \quad T(z) = z\omega'(T(z)), \quad U(z) = z\omega(T) - \frac{T(z)^2}{2}$$

**Point of saddle coalescence.** Graphs with $n$ vertices, $m$ edges, $m = \alpha n(1 - \mu n^{-1/3})$, $\mu \to \infty$,

$$\widehat{z}\frac{\omega''(\widehat{z})}{\omega'(\widehat{z})} = 1, \quad \widehat{z}\frac{\omega'(\widehat{z})}{\omega(\widehat{z})} = 2\alpha.$$

# 2-SAT contradictory component

[Kim '2008], [D. '2019+]

**Point of saddle coalescence.** 2-CNF with $n$ variables, $m$ clauses, $m = n(1 - \mu n^{-1/3})$, $\mu \to \infty$:

$$\mathbb{P}\left(F_{n,m} \text{ is satisfiable}\right) \sim 1 - \frac{1}{16|\mu|^3}$$

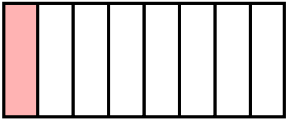
**Theorem.** [D. '2019] Cubic contradictory cores of excess $r$ appear with probability $\Theta(|\mu|^{-3r})$.

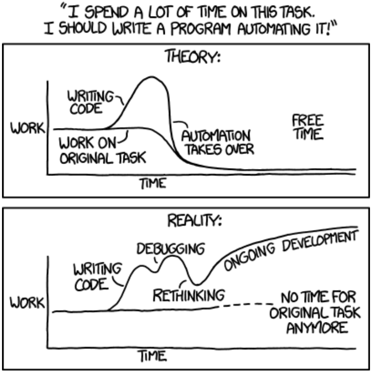
**Open question.** [De Panafieu, D., Ravelomanana '2019]
Digraphs with $n$ vertices, $m$ oriented edges, $m = n(1 - \mu n^{-1/3})$, $\mu \to \infty$:

$$\mathbb{P}\left(\substack{\text{every strong component is} \\ \text{an isolated vertex or a cycle}}\right) \sim 1 - \frac{C}{|\mu|^3}, \quad C = ?$$

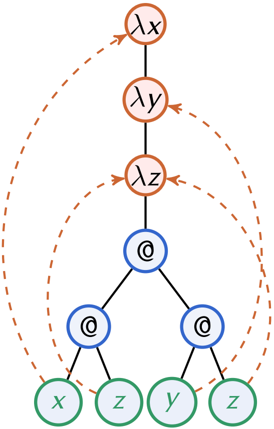
# Part II. Interdisciplinary image

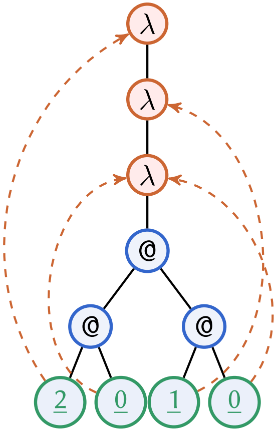
Level I. Software verification





# Lambda terms




A tree-like representation of a *closed* lambda term
$\lambda x.\lambda y.\lambda z.(xz)(yz)$

# Lambda terms

- Church–Turing thesis

  Lambda terms ⇔ Turing machines

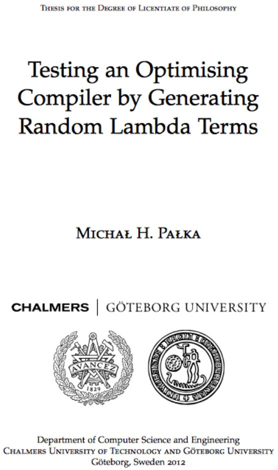
  Beta-reduction step ⇔ Crank turn

- Curry–Howard correspondence

  Typed lambda terms ⇔ Proofs in Hilbert-style deduction logic

  Types ⇔ Statements

# Michal Pałka's thesis


Thesis for the Degree of Licentiate of Philosophy

Testing an Optimising Compiler by Generating Random Lambda Terms

Michał H. Pałka

CHALMERS | GÖTEBORG UNIVERSITY

Department of Computer Science and Engineering
Chalmers University of Technology and Göteborg University
Göteborg, Sweden 2012

- CSmith ['2011]: generation of random C programs, 325 previously unreported bugs in GCC, LLVM, other C compilers and CompCert (with a formally verified core)
- Michal Pałka et al. ['2012]: Eight failures and four bugs detected, reported and fixed (GHC optimising compiler)

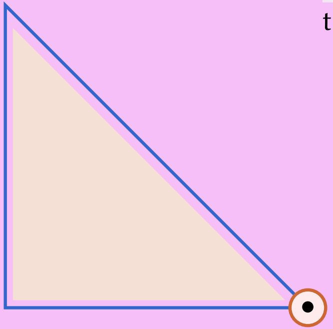
# Statistical properties of closed lambda terms

[Bendkowski, Bodini, D. '2019]

| Parameter | Mean, $\sim$ | | Distribution | |
|---|---|---|---|---|
| | **plain** | **closed** | **plain** | **closed** |
| Variables | $0.307n$ | | Normal | |
| Abstractions | $0.258n$ | | Normal | |
| Successors | $0.129n$ | | Normal | |
| Redexes | $0.091n$ | | Normal | |
| Index value | 0.420 | | Geometric | |
| Redex search time | 6.222 | 6.054 | Discrete | Discrete |
| Head abstractions | 0.420 | 1.447 | Geometric | Discrete |
| $m$-openness | 2.019 | 0 | Discrete | trivial |
| Free variables | 5.722 | 0 | Discrete | trivial |
| Unary height profile | $0.122\sqrt{n}$ | | Rayleigh | |
| Natural height profile | $0.412\sqrt{n}$ | | Rayleigh | |

# Random generation of closed lambda terms

[Bendkowski, Bodini, D. '2019]



**Theorem.** Polynomial algorithm for tuning with given

- number of atomic nodes of distinguished colors
- number of **redexes** (i.e. patterns necessary to perform a computation step in lambda calculus)
- number of head abstractions
- number of closed subterms
- number of any tree-like patterns

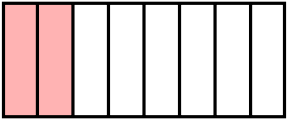
Having *corner cases* instead of *uniform distribution* is necessary.

# Multiparametric Tuner Implementation

[Bendkowski, Bodini, D.]


```python
>>> from paganini import *
>>> spec = Specification()
>>> z, u, M = Variable(1000), Variable(200), Variable()
>>> spec.add(M, z + u * z * M + z * M ** 2)
>>> spec.run_tuner(M)
```

Level II. Data sciences



# Networks

**Definition.** A *network* is a graph or directed graph, formed by *real-world data.*

**Differences exhibited** (networks versus Erdős–Rényi graphs)

- Density
- Degree distribution
- Diameter, shortest path
- Connectivity, clustering
- Spectral properties
- other metrics

# Degree constraints approach

[De Panafieu, Ramos], [D., Ravelomanana]

Set of degree constraints

$$\Delta = \{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \ldots\} \quad \Rightarrow \quad \omega(z) = \sum_{d \in \Delta} \frac{z^d}{d!}$$

- Erdős–Rényi graphs: Poisson degree distribution
- Scale-free network: $\mathbb{P}(k) \sim k^{-\gamma}$

**Theorem.** [D., Ravelomanana]
Phase transition is shifted with "weighted" allowed degrees

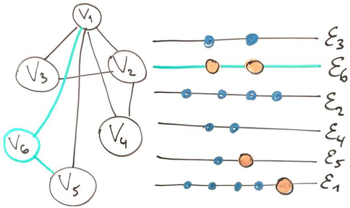
$$\omega(z) = \sum_{d \in \Delta} c_d z^d$$

**Open question.** Can we mimic the behaviour of scale-free network using a weighted EGF for set degree constraints?

# Bianconi–Barabási approach

## Bose–Einstein condensation in evolving networks

| **Bose gas** | **network evolution** |
|---|---|
| diluted gas at low temperature | evolving network |
| temperature | temperature |
| energy | energy |
| particle | half-edge |
| number of energy levels | $\leqslant$ number of nodes |
| Bose–Einstein condensation | topological phase transition |



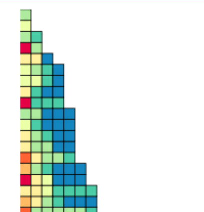

In this model, the number of particles on the energy level $\varepsilon$ follows the Bose statistics $n(\varepsilon) = \frac{1}{e^{\beta(\varepsilon-\mu)}-1}$ which also represents the number of edges linking to nodes with energy $\varepsilon$.

# Weighted partitions as BEC in quantum harmonic oscillator

[Vershik], [Bernstein, Fahrbach, Randall], [Bendkowski, Bodini, D.]

**Coloured partitions $\leftrightarrow$ d-dimensional quantum oscillator**

| Weighted partition | Random particle assembly |
|---|---|
| Sum of numbers | Total energy |
| Number of colours | Dimension ($d$) |
| Row of Young table | Particle |
| Number of rows | Number of particles |
| Number of squares in the row | Energy of a particle ($\lambda$) |
| Partition limit shape | Bose–Einstein condensation |
| $\binom{d+\lambda-1}{\lambda}$ | Number of particle states |



**Theorem.** There exists a polynomial algorithm to generate random assemblies with given numbers of colours $(n_1, n_2, \ldots, n_d)$.

# Combinatorial learning

[D. '2020+]

**Maximum likelihood approach.** Data $X_1, \ldots, X_n$ is i.i.d, and follows $\mathbb{P}_\theta$.

- Parameter estimation: find $\theta$
- Hypothesis testing
- Nonlinear regression: $y = f(x) + \varepsilon$
- Model selection, etc

**Combinatorial learning.** Data $X_1, \ldots, X_n$ are (projections of) parameters of combinatorial objects generated from Boltzmann distribution $\mathbb{P}_z$.

- Object size
- Number of leaves in a tree, height, etc
- Patterns in context-free grammars
- Basically, whatever is markable

# Combinatorial learning

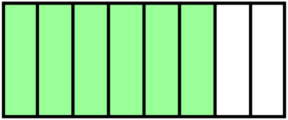
[D. '2020+]

**Hidden parameter estimation.** (Case study)
Objects are sampled from $\mathbb{P}_{z,u}$. We observe only the size parameter corresponding to $z$. Estimate $(z, u)$.

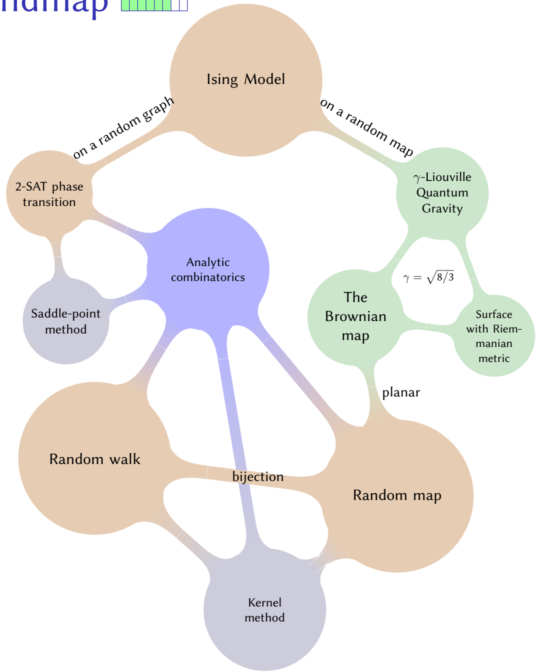
**Lemma.** In case of general dimension, this problem is #P-complete.

**Proposition.** Boltzmann relaxation has a polynomial solution, using the tuning algorithm from [Bendkowski, Bodini, D.]

Level III. Mathematical physics



# Ising model mindmap



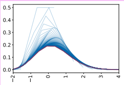
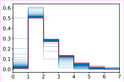
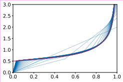
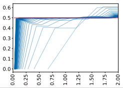
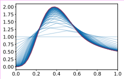
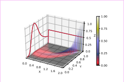
# Connections with mathematical physics

- 2-SAT phase transition in terms of Ising model
- Maps, ribbon graphs and models of 2-d quantum gravity
- Weighted partitions and Bose–Einstein condensate

# Parameters in random maps regardless of genus

[Bodini, Courtiel, D., Hwang]

| Statistics | Differential equation | Mean | Limit law |
|---|---|---|---|
| leaves | $L = v + (2-u)zL + zL^2 + 2z^2\partial_z L + z(1-v)\partial_v L$ | 1 | Poisson(1) |
| root isthmic parts | $C = 1 + zC + vzC\|_{v=1}C + 2z^2\partial_z C$ | 2 | Geometric($\frac{1}{2}$) |
| vertices | $X = v + zX + zX^2 + 2z^2\partial_z X$ | $\log n$ | $\mathcal{N}(\log n, \log n)$ |
| loops | $Y = v + vzY + vzY\|_{v=1}Y + 2vz^2\partial_z Y + v^2z(vw-1)\partial_v Y$ | $\frac{1}{2}n$ | A new law* |
| root edges | $E = 1 + vzE + vzE\|_{v=1}E + 2vz^2\partial_z E$ | $\frac{2}{3}n$ | Beta(1, $\frac{1}{2}$) |
| root degree | $D = 1 + v^2zD + vzD\|_{v=1}D + 2vz^2\partial_z D - v^2(1-v)z\partial_v D$ | $n$ | Uniform[0, 2] |

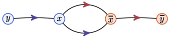
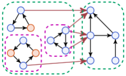
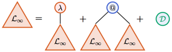
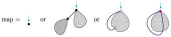







# Part III. Summary

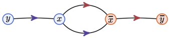
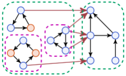
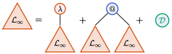
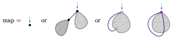
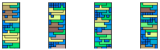
# Crossroad of tools and applications

| | $\lambda$-calculus | networks | mathematical statistics | other |
|---|---|---|---|---|
| $F = \Phi(F, z)$ | closed $\lambda$-terms | | | |
| $z\frac{F'(z)}{F(z)} = n$ | software testing | Bianconi–Barabasi | combinatorial learning | tilings, RNA, queues |
| $P(F, F'_x, F'_y) = 0$ | linear $\lambda$-terms | | | random maps |
| $\int_{\mathbb{R}} e^{-ix^3 - ixy} dx$ | height profile | graphs with degree constraints | | 2-SAT Witten's conjecture |

# Preprints and publications

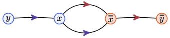
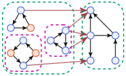
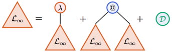
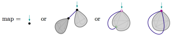
| Paper | | Tools |
|---|---|---|
|  | Shifting the phase transition threshold for random graphs using degree constraints | Combinatorial decomposition + saddle point |
|  | The birth of the contradictory component in random 2-SAT | Inclusion-exclusion + saddle point |
|  | Symbolic method and directed graph enumeration | Combinatorial decomposition + Inclusion-exclusion |
|  | Statistical properties of lambda terms | Infinite-dimensional Drmota–Lalley–Woods Theorem |
|  | Asymptotic distribution of parameters in random maps | Disconnecting transform + linearisation of PDE |
|  | Polynomial tuning of multiparametric combinatorial samplers | Convex optimisation |

# Preprints and publications

| Paper | | Tools |
|---|---|---|
|  | Shifting the phase transition threshold for random graphs using degree constraints | Combinatorial decomposition + saddle point |
|  | The birth of the contradictory component in random 2-SAT | Inclusion-exclusion + saddle point |
|  | Symbolic method and directed graph enumeration | Combinatorial decomposition + Inclusion-exclusion |
|  | Statistical properties of lambda terms | Infinite-dimensional Drmota–Lalley–Woods Theorem |
|   | Asymptotic distribution of parameters in random maps | Disconnecting transform + linearisation of PDE |
|  | Polynomial tuning of multiparametric combinatorial samplers | Convex optimisation |

- Peer-reviewed conference proceedings: 4 papers
- Accepted to a peer-reviewed journal: 1 paper ([70 pages, E-JC])

# Preprints and publications

| Paper | | Tools |
|---|---|---|
|  | Shifting the phase transition threshold for random graphs using degree constraints | Combinatorial decomposition + saddle point |
|  | The birth of the contradictory component in random 2-SAT | Inclusion-exclusion + saddle point |
|  | Symbolic method and directed graph enumeration | Combinatorial decomposition + Inclusion-exclusion |
|  | Statistical properties of lambda terms | Infinite-dimensional Drmota–Lalley–Woods Theorem |
|  | Asymptotic distribution of parameters in random maps | Disconnecting transform + linearisation of PDE |
|  | Polynomial tuning of multiparametric combinatorial samplers | Convex optimisation |

Thank you.