From bipartiteness to 2-SAT: the enumerative approach to random graphs

Vlady Ravelomanana

IRIF – University of Paris Cité.
vlad@irif.fr

Workshop Random Graphs'23 – TU Dortmund

Outline of the talk

- Introduction & motivations.
- Some tractable graph phase transitions.
- 2-SAT with generating functions.
- Conclusion and perspectives.

Main motivations

In Computer Science, two **core** problems (cf. **[Garey, Johnson'79]**).

- Decision problems: write an algorithm such that

$$\begin{cases} \textbf{INPUT :} \text{ an instance } \mathcal{I} \text{ and a property } \mathcal{P} \\ \textbf{OUTPUT :} \text{ YES (resp. NO) if } \mathcal{I} \text{ satisfies (resp. does not satisfy) } \mathcal{P} \end{cases}$$

- Optimization problems: find the **best solution** from all feasible solutions.

Introduction

Main motivations

In Computer Science, two **core** problems (cf. **[Garey, Johnson'79]**).

- Decision problems: write an algorithm such that

$$\begin{cases} \textbf{INPUT} : \text{an instance } \mathcal{I} \text{ and a property } \mathcal{P} \\ \textbf{OUTPUT} : \text{YES (resp. NO) if } \mathcal{I} \text{ satisfies (resp. does not satisfy) } \mathcal{P} \end{cases}$$

- Optimization problems: find the **best solution** from all feasible solutions.

Other communities: Probability/Combinatorics/Physics

C: configuration of the system, **E:** energy function, **T:** temperature, **Z:** norm. constant **[Baldassi, Braunstein, Ramezanpour, Zecchina'09]**

$$\mathbb{P}(C) = \frac{1}{Z} e^{-E(C)/T}$$

- if $T = \infty$ then all configurations C are equiprobable, the system is **disordered**
- if $T = 0$ $\mathbb{P}(C)$ is "concentrated on the minimum energy function" (the *ground state*). Optimizing $\to$ finding a zero energy ground state of E.

The 3-SAT example

- Formula built with n variables ($2n$ literals) $x_1,\ x_2,\ \cdots,\ x_n$ and $\bar{x}_1,\ \bar{x}_2,\ \cdots,\ \bar{x}_n$.
- M clauses implying 3 different variables per clause (constraints on the literals)

Question?

Can we find assignment(s) of the x_i satisfying (SAT) all the clauses? Ex:

$$(x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee \bar{x}_3) \wedge (x_1 \vee \bar{x}_2 \vee x_3) \wedge (x_1 \vee \bar{x}_2 \vee \bar{x}_3)$$
$$\wedge$$
$$(\bar{x}_1 \vee x_2 \vee x_3) \wedge (\bar{x}_1 \vee x_2 \vee \bar{x}_3 \wedge (\bar{x}_1 \vee \bar{x}_2 \vee x_3 \wedge (\bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3)$$

Translating CS-langage to Physics-langage: the k-SAT example

Computer Science	**Statistical Physics**
Boolean variable $x \in \{\text{True, False}\}$	Ising spin $s \in \{\text{spin up}(+1), \text{Spin down}(-1)\}$
Clauses	Couplings and fields acting on spins
Number of clauses violated	Energy E of spins configuration
2-SAT : $(x \text{ or } \bar{y})$ and $(\bar{x} \text{ or } z)$	$E = \frac{1}{4}(1-s_x)(1+s_y) + \frac{1}{4}(1+s_x)(1-s_z)$
3-SAT : $(x \text{ or } \bar{y} \text{ or } z)$	$E = \frac{1}{8}(1-s_x)(1+s_y)(1-s_z)$
The problem is SAT	Ground state energy $= 0$
The problem is UNSAT	Ground state energy > 0

cf. works of **[Mézard]**, **[Monasson]**, **[Montanari]**, **[Ricci-Tersenghi]**, **[Parisi]**.

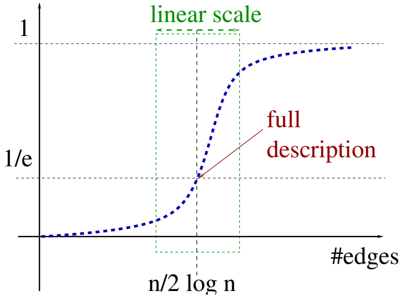
The connectivity phase transition: a typical theorem

$G(n, M)$: graph with M edges uniformly drawn from $\binom{n}{2}$ edges.

From "*On Random Graphs I*" [ERDŐS, RÉNYI 59]

Theorem. As $M = \frac{1}{2}n\Big(\log n + c(n)\Big)$

$$\lim_{n\to+\infty} \texttt{Proba}\,[G(n,M) \text{ is connected}] = \begin{cases} 0 \text{ as } c(n) \to -\infty \\ e^{-e^{-c}} \text{ as } c(n) \to c \text{ (constant)} \\ 1 \text{ as } c(n) \to +\infty. \end{cases}$$



Main research directions

- Random *k*-**SAT** formulas are subject to phase transition phenomena **[Friedgut,Bourgain'99]**
- Main research tasks include:
  1. **Localization** of the threshold (ex: 3-**SAT** $4.2\cdots$ cf. **[Coja-Oghlan, Panagiotou '16]**. *k*-**XORSAT** cf. **[Coja-Oghlan, Kang, Krieg, Rolvien'23++]**. *k*-SAT**[Ding, Sly, Sun '22]**.
  2. Nature of the transition: **sharp/coarse**. **[Creignou, Daudé' 09]**
  3. **Scaling window** (e.g. 2-SAT **[Bollobàs, Borgs, Kim, Wilson'01]**) and/or (if possible) details inside the window of **transition**.
  4. **Algorithmic complexity** of the decision problem (e.g. "2-SAT is in P" **[Tarjan '79]**).

Main research directions

- Random *k*-**SAT** formulas are subject to phase transition phenomena **[Friedgut,Bourgain'99]**
- Main research tasks include:
  1. **Localization** of the threshold (ex: 3-**SAT** $4.2\cdots$ cf. **[Coja-Oghlan, Panagiotou '16]**. *k*-**XORSAT** cf. **[Coja-Oghlan, Kang, Krieg, Rolvien'23++]**. *k*-**SAT** **[Ding, Sly, Sun '22]**.
  2. Nature of the transition: **sharp/coarse**. **[Creignou, Daudé' 09]**
  3. **Scaling window** (e.g. 2-SAT **[Bollobàs, Borgs, Kim, Wilson'01]**) and/or (if possible) details inside the **window of transition**.
  4. **Algorithmic complexity** of the decision problem (e.g. "2-SAT is in P" **[Tarjan '79]**). **Only tractable problems today!**

Outline of the talk

- Introduction & motivations.
- Some tractable graph phase transitions.
- 2-SAT with generating functions.
- Conclusion and perspectives.

- **2-COL (bipartiteness):** is a graph **2-COLORABLE**?
- **2-XORSAT :** is the formula

$$\begin{cases} x_1 \oplus x_2 = 1 \\ x_2 \oplus x_3 = 0 \\ \dots \\ x_i \oplus x_j = \varepsilon \in \{0,\ 1\} \\ \dots \end{cases} \qquad \textbf{SAT}?$$

- **Is a random directed graph acyclic?**
- **2-SAT:** is the formula

$$\begin{cases} \dots \\ r \vee s \\ \dots \end{cases} \qquad \textbf{SAT}?$$

(where r and $s \in \{x_1, \cdots, x_n\} \bigcup \{\bar{x}_1, \cdots, \bar{x}_n\}$)

- $\dots$

- **General form :**

$$A.X = \varepsilon$$

where A has m rows and ε a m-dimensional $0/1$ vector.

- **Distribution :** uniform. Pick m clauses of the form $x_i \oplus x_j \in \{0, 1\}$ from the set of $n(n-1)$ clauses.
- **Underlying structures :** **`graphs with weighted edges`**.

- **General form :**

$$A.X = \varepsilon$$

  where A has m rows and ε a m-dimensional $0/1$ vector.

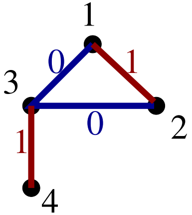
- **Distribution :** uniform. Pick m clauses of the form $x_i \oplus x_j \in \{0,\ 1\}$ from the set of $n(n-1)$ clauses.
- **Underlying structures :** **graphs with weighted edges**.

SAT characterization

SAT iff no elementary cycle of odd weight.

SAT iff no elementary cycle of odd weight

$$\begin{cases} x_1 \oplus x_2 = 1 \\ x_2 \oplus x_3 = 0 \\ x_1 \oplus x_3 = 0 \\ x_3 \oplus x_4 = 1 \end{cases}$$



Proof.

- UNSAT $\Longleftarrow$ Fix a cycle of odd weight, ...
- SAT $\Longleftarrow$ No cycles of odd weight, a DFS affectation based proof.

A basic scheme

- Enumeration of “SAT”-graphs (graphs without cycles of odd weight by means of generating functions).

- Use the results to compute

$$\texttt{Proba SAT} = \frac{\sharp\texttt{config. without cycles of odd weight}}{\sharp\texttt{total configurations}}.$$

Combinatorics

A basic scheme

- Enumeration of "SAT"-graphs (graphs without cycles of odd weight by means of generating functions).

- Use the results to compute

$$\texttt{Proba SAT} = \frac{\sharp\texttt{config. without cycles of odd weight}}{\sharp\texttt{total configurations}}.$$

Works on 2-XORSAT

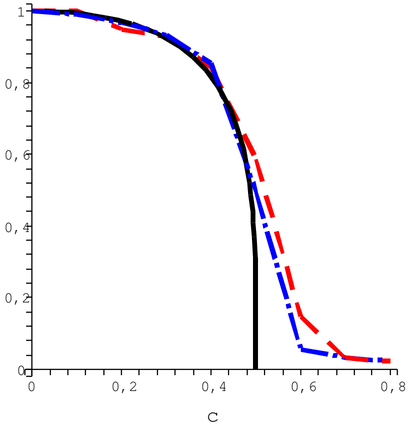
- Using statistical physics methods, **[Monasson '07] inferred** that

$$\lim_{n\to+\infty} n^{\texttt{critical exponent}} \times \mathbb{P}\left[\text{2-XORSAT}\left(n, \frac{n}{2}\right)\right] = \mathbf{O(1)}$$

where **critical exponent** must be $\frac{1}{12}$.

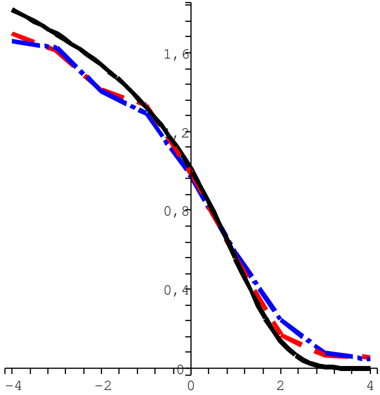
- In **[Daudé, R. '11]**, enumerative approaches are used to show that **YES** critical exponent is $1/12$ and the **O(1)** are explicitly computed.

Taste of our results : the whole window



p(n, cn) $\stackrel{\text{def}}{=}$ `Proba` **[2** – XOR with **n** variables **, cn** clauses **]** is **SAT**
for $n = 1000$, $n = 2000$ and the **theoretical** function : $\mathbf{e^{c/2}(1-2c)^{1/4}}$.

Taste of our results: rescaling the critical window



Rescaling at the point "zero", i.e $c = 1/2$: $n = 1000$, $n = 2000$ and $\lim_{n\to\infty} \; n^{1/12} \times \; p(n, n/2 + \mu n^{2/3})$ as a **function of** μ.

Basic refresher of Symbolic Method on EGFs

The **exponential generating function** (EGF) associated to a **labelled class** $\mathcal{A}$ of combinatorial objects is

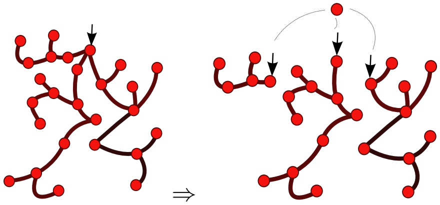
$$A(z) = \sum_{a \in \mathcal{A}} \frac{z^{|a|}}{|a|!} = \sum_{n \geqslant 0} a_n \frac{z^n}{n!}\,.$$

cf **[Flajolet - Sedgewick'09]**

Construction	Notation	Comments	EGFs
Disjoint union	$\mathcal{A} + \mathcal{B}$	disjoint copies of objects from $\mathcal{A}$ and $\mathcal{B}$	$A(z) + B(z)$
Labeled product	$\mathcal{A} \star \mathcal{B}$	ordered pairs of copies one from $\mathcal{A}$ and one from $\mathcal{B}$	$A(z)B(z)$
Sequence	$\mathrm{Seq}(\mathcal{A})$	sequences of objects from $\mathcal{A}$	$\frac{1}{1-A(z)}$
Set	$\mathrm{Set}(\mathcal{A})$	set of objects from $\mathcal{A}$	$e^{A(z)}$
Cycle	$\mathrm{Cyc}(\mathcal{A})$	cycles of k objects from $\mathcal{A}$	$\log \frac{1}{1-A(z)}$

Trees

We apply the previous grammar to count *rooted* trees



$$\mathcal{T} = \bullet \times \mathrm{Set}(\mathcal{T}) \to T(x) = xe^{T(x)}$$

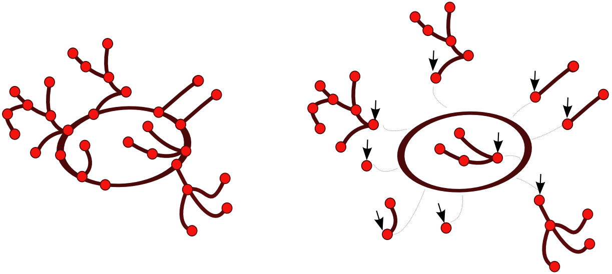
To forget the root, we just integrate: $(xU'(x) = T(x))$

$$\int_0^x \frac{T(s)}{s} ds = \quad \begin{array}{c} T(s) = u \\ T'(s)\, ds = du \end{array} \quad = \int_{T(0)}^{T(x)} 1 - u\, du = T(x) - \frac{1}{2} T(x)^2$$

and the general version

$$e^{U(x)} = e^{T(x)} e^{-\frac{1}{2}T(x)^2}$$

Unicyclic graphs



$$\mathcal{V} = \bigcirc_{\geq 3}(\mathcal{T}) \to V(x) = \sum_{n=3}^{\infty} \frac{1}{2} \frac{(n-1)!}{n!} (T(x))^n$$

We can write $V(x)$ in a compact way:

$$\frac{1}{2} \quad -\log\left(1 - T(x)\right) - T(x) - \frac{T(x)^2}{2} \quad \to e^{V(x)} = \frac{e^{-T(x)/2 - T(x)^2/4}}{\sqrt{1 - T(x)}}.$$

Generating functions (ctd)

Connected graphs without cycles of odd weight (2-XORSAT)

We want to **enumerate**

- **labelled connected graphs** (nodes labelled in $[1, n]$) with edges of weight in $\{0, 1\}$ (cf. **[Harary, Palmer '73]**)
- according to two parameters
  1. number of vertices n
  2. number of edges $n + k$ (k is the excess of the component, $k \geqslant -1$).
- SAT = **those without cycles of odd weight.**

Let

$$C_k(z) = \sum_{k>0} c_{n,n+k} \frac{z^n}{n!} .$$

Generating functions (ctd)

Connected graphs without cycles of odd weight (2-XORSAT)

We want to **enumerate**

- **labelled connected graphs** (nodes labelled in $[1, n]$) with edges of weight in $\{0, 1\}$ (cf. **[Harary, Palmer '73]**)
- according to two parameters
  1. number of vertices n
  2. number of edges $n + k$ (k is the excess of the component, $k \geqslant -1$).
- SAT = **those without cycles of odd weight.**

Let

$$C_k(z) = \sum_{k \geq 0} c_{n,n+k} \frac{z^n}{n!} .$$

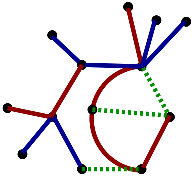
Th. [Daudé, R. 11]

Let Wright(z) be the number of unweighted connected graphs of excess k (**[Wright '77]**). Then

$$C_k(z) = \frac{1}{2} \, \mathtt{ConnectedGraphs}_k(2z) .$$

Proof.

- Recall *SAT* iff **`no cycles of odd weight`**
- Consider a “SAT” connected component g
- Pick a spanning tree of g with n nodes:



Proof.

- Recall *SAT* iff **no cycles of odd weight**
- Consider a "SAT" connected component g
- Pick a spanning tree of g with n nodes:

- The edges of the spanning tree can be weighted in 2^{n-1} ways. By **[Diestel'00]**, the weights of the other edges are **determined**:

$$C_k(z) = \sum_{k^0} 2^{n-1} \underbrace{\texttt{connected graphs}(n,\, n+k)}_{\texttt{[Wright '77]}} \frac{z^n}{n!} .$$

Main ideas: n variables, m clauses

$$\texttt{Proba SAT} = n! \times \frac{[z^n]\textbf{SAT-GF}(z)}{\binom{n(n-1)}{m}}.$$

As the number of clauses m increases, random formulas behave as random graphs with **gazeous**, **liquid** and **hardening** phases.

- **Sub-critical phase (gaz) :** forest of trees and set of unicyclic components.
- **Critical phase (liquid) :** forest of trees, set of unicyclic components and few multicyclic components.
- **Super-critical phase (hardening) :** forest of trees, set of unicyclic components and a single (baby) giant component.

The Random 2-XORSAT Transition

Th. [Daudé, R. '11]

The probability that a random formula with **n** variables and **m** clauses is **SAT** satisfies the following :

(i) Sub-critical phase. As $0 < n - 2m \ll n^{2/3}$

$$\mathbb{P}(n,\ m) = e^{m/2n}\left(1 - 2\frac{m}{n}\right)^{1/4} + O\left(\frac{n^2}{(n-2m)^3}\right)\ .$$

(ii) Critical phase. As $m = \frac{n}{2} + \mu n^{2/3}$, μ fixed real

$$\lim_{n\to+\infty} n^{1/12}\mathbb{P}\left(n,\ \frac{n}{2}(1+\mu n^{-1/3})\right) = \Psi(\mu),$$

where Ψ can be explicitly expressed (in terms of the Airy function).

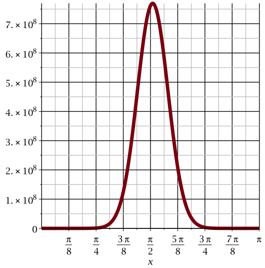
(iii) Super-critical phase. As $m = \frac{n}{2} + \mu n^{2/3}$ with $\mu = o(n^{1/12})$

$$\mathbb{P}\left(n,\ \frac{n}{2}(1+\mu n^{-1/3})\right) = \mathrm{Poly}(n,\mu)e^{-\frac{\mu^3}{6}}\ .$$

For the Asymptotics: The classical saddle-point method

The real case looks like (Laplace's method):

$$R(n) = \int_{x_1}^{x_2} g(x) e^{nh(x)} dx\,.$$



$$\int_0^{\pi} x e^{20 \sin x} dx = 4.29 \times 10^8\,.$$

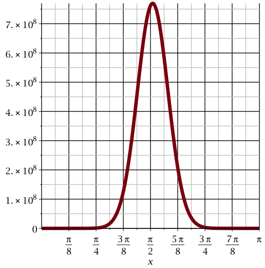
$$\int_0^{\pi} x e^{20(1-\frac{1}{2}(x-\pi/2)^2)} dx = 4.27 \times 10^8\,.$$

Normal saddle point x_0 s.t $h'(x_0) = 0$ and $h''(x_0) \neq 0$.

For the Asymptotics: The classical saddle-point method

The real case looks like (Laplace's method):

$$R(n) = \int_{x_1}^{x_2} g(x) e^{nh(x)} dx \, .$$



$$\int_0^{\pi} x e^{20 \sin x} dx = 4.29 \times 10^8 \, .$$
$$\int_0^{\pi} x e^{20(1-\frac{1}{2}(x-\pi/2)^2)} dx = 4.27 \times 10^8 \, .$$

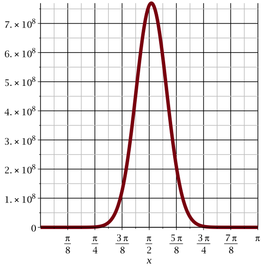
Normal saddle point x_0 s.t $h'(x_0) = 0$ and $h''(x_0) \neq 0$.

What if $h''(x_0) = 0$?

The classical saddle-point method

The real case looks like (Laplace's method):

$$R(n) = \int g(x)e^{nh(x)}dx\, g(x_0) \qquad \sim \qquad \sqrt{\frac{2\pi}{-h''(x_0)n}}\, e^{nh(x_0)}$$



$$\int_0^{\pi} xe^{20\sin x}dx = 4.29 \times 10^8 .$$

$$\int_0^{\pi} xe^{20(1-\frac{1}{2}(x-\pi/2)^2)}dx = 4.27 \times 10^8 .$$

Normal saddle point x_0 s.t $\mathbf{h'(x_0) = 0}$ and $\underline{\underline{\mathbf{h''(x_0) \neq 0}}}$.

Asymptotics: The case with $h''(z_0) = 0$ (Airy instead of Gauss)

Consider

$$\Psi''(z) - z\Psi(z) = 0 \qquad \textbf{(Airy equation)}\,.$$

Look for a solution of the form

$$\Psi(z) = \int_\Gamma e^{tz}\Phi(t)dt \textbf{ with } \Gamma \textbf{ in the complex } t - \textbf{plane}$$

(Φ with no singularities on Γ.)

$$\Psi''(z) = \int_\Gamma t^2 e^{tz}\Phi(t)dt \qquad \textbf{and} \qquad z\Psi(z) = [e^{tz}\Phi(t)]_{\delta\Gamma} - \int_\Gamma e^{tz}\Phi'(t)dt\,.$$

We have then

$$t^2\Phi(t) + \Phi'(t) = 0 \textbf{ leading to solution of the form } \Psi(z) = c\int_\Gamma \exp\left(tz - \frac{t^3}{3}\right)dt$$

(**plus** the condition $e^{tz}\Phi(t) = 0$ on the boundary of the contour Γ).

Proof of the Theorem : the sub-critical phase

1. As $0 < n - 2m \ll n^{2/3}$,the probability that the underlying graph has no multicyclic component is

$$1 - O\left(\frac{n^2}{(n-2m)^3}\right) .$$

Proof of the Theorem : the sub-critical phase

1. As $0 < n - 2m \ll n^{2/3}$, the probability that the underlying graph has no multicyclic component is
$$1 - O\left(\frac{n^2}{(n-2m)^3}\right) .$$

2. Then, the probability that the random formula is SAT is close to
$$\frac{n!}{\binom{n(n-1)}{m}}[z^n]\frac{(\text{unrooted trees}(z))^{n-m}}{(n-m)!} \times \text{set of even weighted unicycles}(z) .$$

Proof of the Theorem : the sub-critical phase

1. As $0 < n - 2m \ll n^{2/3}$,the probability that the underlying graph has no multicyclic component is
$$1 - O\left(\frac{n^2}{(n-2m)^3}\right) .$$

2. Then, the probability that the random formula is SAT is close to
$$\frac{n!}{\binom{n(n-1)}{m}} [z^n] \frac{(\text{unrooted trees}(z))^{n-m}}{(n-m)!} \times \text{set of even weighted unicycles}(z) .$$
That is $\text{Stirling}(n, m) \times \text{Cauchy}(z)$ with
$$\text{Cauchy}(z) = \frac{1}{2\pi i} \oint \left(\frac{T(2z)}{2} - \frac{T(2z)^2}{4}\right)^{n-m} \frac{e^{-T(2z)/4 - T(2z)^2/8}}{(1-T(2z))^{1/4}} \frac{dz}{z^{n+1}}$$
and $T = -LambertW(-z) = \sum_{n>0} n^{n-1} \frac{z^n}{n!}$.

3. Lagrangian subs. $u = T(2z)$ leads to integral of the form $\oint g(u) \exp(nh(u)) du$ with $h(u) = u - \frac{m}{n} \log u + (1 - m/n) \log(2-u)$ so that $h'(2m/n) = 0$ and $h''(2m/n) > 0$ $\rightarrow$ **classical saddle point method** applies on circular path $|z| = 2m/n$.

Proof of the Theorem : the critical phase

Some **multicyclic components** can appear and the general formula looks like

1. $$\mathrm{coeff(n, m,\ r)} \times \frac{1}{2\pi i} \oint \left(\frac{T(2z)}{2} - \frac{T(2z)^2}{4} \right)^{n-m+r} \frac{e^{-T(2z)/4 - T(2z)^2/8}}{(1 - T(2z))^{1/4+3r}} \frac{dz}{z^{n+1}}$$

2. Again
$$\mathrm{Stirling(n, m, r)} \times \frac{1}{2\pi i} \oint g_r(u) e^{nh(u)} du$$

3. $h(u) = u - \frac{m}{n} \log u + (1 - m/n) \log (2 - u)$ but this time there are 2 **saddle-points**:
$$u_0 = 2m/n = 1 + 2\,\mu\, n^{-1/3} \text{ and } u_1 = 1\,.$$
(observe $u_0 \sim u_1$ as $n \gg 1$.)

4. Moreover, $h(1) = h'(1) = h''(1) = 0$ so that the **usual "Gaussian" approximation does not hold**.

The critical phase via Airy function

Integral representation on the complex plane

The Airy function is given by

$$\mathrm{Ai}(z) = \frac{1}{2\pi i}\int_{\mathcal{C}} \exp\left(\frac{t^3}{3} - z\,t\right) dt \; = \frac{\sqrt{3}}{2\pi}\int_0^{\infty} \exp\left(-\frac{t^3}{3} - \frac{z^3}{3t^3}\right) dt$$

where the integral is over a path $\mathcal{C}$ starting at the point at infinity with argument $-\frac{\pi}{3}$ and ending at the point at infinity with argument $+\frac{\pi}{3}$

References

- Books: **[Wong'89]**, **[Olver'97]**.
- Random graphs (analytic combinatorics approaches): **[Flajolet, Knuth, Pittel'89]**, **[Janson, Knuth, Łuczak, Pittel'93]**.
- Analytic combinatorics with several variables: **[Pemantle, Wilson, Melczer'23]**.

Proof of the critical phase of the Th.

Let $p_r(n, m) =$ Proba. to have SAT-graph of excess r.

Thus, $p(n, m) = \sum_{r\geqslant 0} p_r(n, m)$ is the probability of a random formula being SAT.

The proof of part (ii) of the Theorem (and confirmation of **[Monasson '07]**) is completed as

1. For fixed r, we compute (by means of the Airy stuff)

$$n^{1/12} \times p_r(n, m) \sim c_r A(3r + 1/4, \mu)$$

2. and there exist R, C and ε such that for $r \geqslant R$ and all n:

$$n^{1/12} p_r(n, m) \leqslant Ce^{-\varepsilon r}$$

The proof is completed using dominated convergence theorem.

Proof of the critical phase of the Th.

Let $p_r(n, m) =$ Proba. to have SAT-graph of excess r.

Thus, $p(n, m) = \sum_{r \geqslant 0} p_r(n, m)$ is the probability of a random formula being SAT.

The proof of part (ii) of the Theorem (and confirmation of **[Monasson '07]**) is completed as

1. For fixed r, we compute (by means of the Airy stuff)

$$n^{1/12} \times p_r(n, m) \sim c_r A(3r + 1/4, \mu)$$

2. and there exist R, C and ε such that for $r \geqslant R$ and all n:

$$n^{1/12} p_r(n, m) \leqslant Ce^{-\varepsilon r}$$

The proof is completed using dominated convergence theorem.

Generalization

The Random 2-XORSAT Problem has been generalized in **[de Panafieu '14]**'s PhD thesis.

Recap. of some tractable phase transitions

- **2-XORSAT :** is a random formula **SAT**? Done.

Recap. of some tractable phase transitions

- **2-XORSAT :** is a random formula **SAT**? Done.
- **2-COL (bipartiteness):** is a random graph **2-COLORABLE**? Done by **[Pittel, Yeum '10]**

Recap. of some tractable phase transitions

- **2-XORSAT :** is a random formula **SAT**? Done.
- **2-COL (bipartiteness):** is a random graph **2-COLORABLE**? Done by **[Pittel, Yeum '10]**
- **Is a random digraph acyclic?** Work by **[Dovgal, de Panafieu, Ralaivaosaona, Rasendrahasina, Wagner '23]**

Recap. of some tractable phase transitions

- **2-XORSAT :** is a random formula **SAT**? Done.
- **2-COL (bipartiteness):** is a random graph **2-COLORABLE**? Done by **[Pittel, Yeum '10]**
- **Is a random digraph acyclic?** Work by **[Dovgal, de Panafieu, Ralaivaosaona, Rasendrahasina, Wagner '23]**
- **2-SAT:** is a random formula like

$$\begin{cases} x_1 \vee \bar{x}_{19} \\ \dots \\ \bar{x}_{27} \vee \bar{x}_{36} \\ \dots \\ \textbf{SAT?} \\ \textbf{Exact enumeration formulas. Done [Dovgal, de Pana} \\ \text{Asymptotics: open problem!} \end{cases}$$

Random digraphs

- Consider random directed graphs
  Critical random graphs $\overrightarrow{D}(n, n)$ with n nodes and n arcs OR $\overrightarrow{D}(n, p = 1/n)$.
- What is known:
  1. **[Karp '90]** $p = c/n$ with $c < 1$ or $c > 1$.
     OUTSIDE the scaling window!

Random digraphs

- Consider random directed graphs
  Critical random graphs $\overrightarrow{D}(n, n)$ with n nodes and n arcs OR $\overrightarrow{D}(n, p = 1/n)$.
- What is known:
  1. **[Karp '90]** $p = c/n$ with $c < 1$ or $c > 1$.
     OUTSIDE the scaling window!
  2. **[Łuczak '90]** $p = c/n$ $c < 1$ or $c > 1$
     small or giant strongly connected components

     See also **[Pittel, Poole '14]** for asymptotic normality).
     OUTSIDE the scaling window!

Random digraphs

- Consider random directed graphs
  Critical random graphs $\overrightarrow{D}(n, n)$ with n nodes and n arcs OR $\overrightarrow{D}(n, p = 1/n)$.
- What is known:
  1. **[Karp '90]** $p = c/n$ with $c < 1$ or $c > 1$.
     OUTSIDE the scaling window!
  2. **[Łuczak '90]** $p = c/n$ $c < 1$ or $c > 1$
     small or giant strongly connected components

     See also **[Pittel, Poole '14]** for asymptotic normality).
     OUTSIDE the scaling window!

Random digraphs

- Consider random directed graphs
  Critical random graphs $\overrightarrow{D}(n, n)$ with n nodes and n arcs OR $\overrightarrow{D}(n, p = 1/n)$.
- What is known:
  1. **[Karp '90]** $p = c/n$ with $c < 1$ or $c > 1$.
     OUTSIDE the scaling window!
  2. **[Łuczak '90]** $p = c/n$ $c < 1$ or $c > 1$
     small or giant strongly connected components

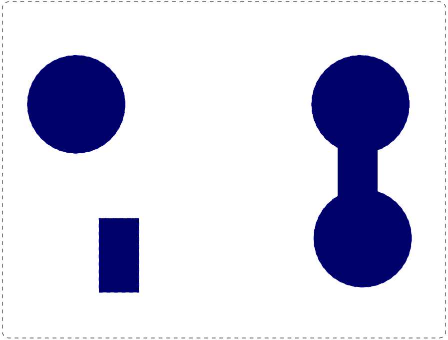
     See also **[Pittel, Poole '14]** for asymptotic normality).
     OUTSIDE the scaling window!

Going inside the scaling window

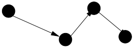
Dovgal *et al* "The birth of the strong components" (2023)
https://arxiv.org/pdf/2009.12127.pdf
Characterize typical strongly connected components of $\overrightarrow{D}(n,\ n + O(n^{2/3}))$.

Intuition.

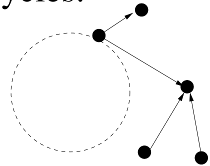
Connected giant component: O(n)



tree–like components:



cycles:



Strongly connected sub–components: $O(n^{1/3})$

Some (very recent) works include

- **[Coulson'22]** Largest connected component inside the window is $O(n^{1/3})$
- **[Blanc-Renaudie'23]** Directed paths also are of order $O(n^{1/3})$ as well as diameter.
- **[Goldschmidt and Stephenson'23]** The scaling limit of a critical random directed graph: they show that strongly connected components (SCC) ranked in decreasing order of size and rescaled by $n^{-1/3}$, converge in distribution to a sequence $(C_1, C_2, \cdots)$ of finite strongly connected directed multigraphs with edge lengths which are either 3-regular or loops.

[Dovgal, de Panafieu, Ralaivaosaona, Rasendrahasina, Wagner'23]

"The birth of the strong components"

Let $c = M/n < 1$. Say digraph is elementary if all strong cc (SCC) are either isolated vertices or directed cycles:

$$\mathbb{P}\left[D(n, M) \text{ is } \text{elementary}\right] \sim exp(-c)(1-c)\,.$$

See also **[Luczak, Seierstad'09]**).
Note that as c tends to 1 such that $c = 1 + \mu n^{-2/3}$, the probability should be *continous* (between the subcritical and the critical phases).

This is very similar to **2-XORSAT:** recall the zoom with a factor $\mathbf{n^{1/12}}$ the *critical exponent* inferred by physicists.

For digraphs, to compute the probability as a function of μ (inside the window), we need to zoom by a factor of order $\mathbf{n^{1/3}}$.

Start with the beautiful enumeration result of **[Robinson'73]**:

$$\mathrm{DAG}(n,\ M) = n![z^n w^M]\,\frac{(1+w)^{\binom{n}{2}}}{\sum_{a\geqslant 0}(1+w)^{-\binom{a}{2}}\frac{(-z)^a}{a!}}$$

and recall the EGF for labelled graphs: $G(z, w) = \sum_{n\geqslant 0}(1+w)^{\binom{n}{2}}\frac{z^n}{n!}$ and

Start with the beautiful enumeration result of **[Robinson'73]**:

$$\mathrm{DAG}(n,\ M) = n![z^n w^M]\,\frac{(1+w)^{\binom{n}{2}}}{\sum_{a\geqslant 0}(1+w)^{-\binom{a}{2}}\frac{(-z)^a}{a!}}$$

and recall the EGF for labelled graphs: $G(z,w) = \sum_{n\geqslant 0}(1+w)^{\binom{n}{2}}\frac{z^n}{n!}$ and its decomposition (set of trees **U,** set of unicycles **V** and complex connected components **Complex$_k$):**

$$G(z,w) = e^{U(wz)/w}e^{V(wz)}\sum_{k\geqslant 0}\mathrm{Complex}_k(wz)w^k\,.$$

From DAG to BIgraphs (2/3)

Observe that in

$$\mathrm{DAG}(n,\ M) = n![z^n w^M]\,\frac{(1+w)^{\binom{n}{2}}}{\sum_{a\geqslant 0}\frac{1}{(1+w)^{\binom{a}{2}}}\frac{(-z)^a}{a!}}\,,$$

we have two occurrences of patterns of $G(z,\ w) = \sum_{n\geqslant 0}(1+w)^{\binom{n}{2}} z^n/n!$.

From DAG to BIgraphs (2/3)

Observe that in

$$\mathrm{DAG}(n,\ M) = n![z^n w^M] \frac{(1+w)^{\binom{n}{2}}}{\sum_{a \geqslant 0} \frac{1}{(1+w)^{\binom{a}{2}}} \frac{(-z)^a}{a!}},$$

we have two occurrences of patterns of $G(z,\ w) = \sum_{n \geqslant 0} (1+w)^{\binom{n}{2}} z^n / n!$. The numerator contains

$$(1+w)^{\binom{n}{2}} = n![z_0^n] G(z_0,\ w)$$

as the denominator is $G(-z,\ \frac{-w}{1+w})$. Rewritting, one obtains

$$\mathrm{DAG}(n,\ M) = (n!)^2 [{z_0}^n {z_1}^n w^M] \frac{G(z_0,\ w)}{G(-z_1,\ -\frac{w}{1+w})}.$$

From DAG to BIgraphs (2/3)

Observe that in

$$\mathrm{DAG}(n,\ M) = n![z^n w^M]\,\frac{(1+w)^{\binom{n}{2}}}{\sum_{a\geqslant 0}\frac{1}{(1+w)^{\binom{a}{2}}}\frac{(-z)^a}{a!}}\,,$$

we have two occurrences of patterns of $G(z,\ w) = \sum_{n\geqslant 0}(1+w)^{\binom{n}{2}} z^n/n!$.
The numerator contains

$$(1+w)^{\binom{n}{2}} = n![z_0^n]G(z_0,\ w)$$

as the denominator is $G(-z,\ \frac{-w}{1+w})$. Rewritting, one obtains

$$\mathrm{DAG}(n,\ M) = (n!)^2[{z_0}^n {z_1}^n w^M]\,\frac{G(z_0,\ w)}{G(-z_1,\ -\frac{w}{1+w})}\,.$$

IMPORTANT

Looks (resp. behaves) like two graphs with **color '0'** and **color '1'** with the same number of vertices in the sparse regime.

In fact,

$$\mathrm{DAG}(n,\ M) = (n!)^2[{z_0}^n {z_1}^n w^M]\, \frac{G(z_0,\ w)}{G(-z_1,\ -\frac{w}{1+w})}$$

after substitutions $(z_0,\ z_1,\ w) \to (\frac{z_0}{y},\ \frac{1+y}{y} z_1, y)$, one gets

$$\mathrm{DAG}(n,\ M) = (n!)^2[{z_0}^n {z_1}^n y^{M-2n}] e^{U(z_0)/y + U(z_1)/y}\ e^{U(z_1)} \tag{1}$$

$$\frac{\sum_{j \geqslant 0} \mathrm{Complex}_j(z_0) y^j}{\sum_{k \geqslant 0} \mathrm{Complex}_k(z_1) \left(-\frac{y}{1+y}\right)^k}\ \frac{e^{V(z_0) - V(z_1)}}{(1+y)^n}\ . \tag{2}$$

From DAG to BIgraphs (3/3)

In fact,

$$\mathrm{DAG}(n,\ M) = (n!)^2 [{z_0}^n {z_1}^n w^M]\, \frac{G(z_0,\ w)}{G(-z_1,\ -\frac{w}{1+w})}$$

after substitutions $(z_0,\ z_1,\ w) \to (\frac{z_0}{y},\ \frac{1+y}{y} z_1, y)$, one gets

$$\mathrm{DAG}(n,\ M) = (n!)^2 [{z_0}^n {z_1}^n y^{M-2n}] e^{U(z_0)/y + U(z_1)/y}\, e^{U(z_1)} \tag{1}$$

$$\frac{\sum_{j\geqslant 0} \mathrm{Complex}_j(z_0) y^j}{\sum_{k\geqslant 0} \mathrm{Complex}_k(z_1) \left(-\frac{y}{1+y}\right)^k}\, \frac{e^{V(z_0) - V(z_1)}}{(1+y)^n}\,. \tag{2}$$

REMARK

Counting the 'y', the excesses and the trees, this can be decomposed into an infinite sum of GFs:

$$(n!)^2 \sum_{t\geqslant 0} [{z_0}^n {z_1}^n] \frac{(U(z_0) + U(z_1))^{2n-m+t}}{(2n-m+t)!}\, \frac{e^{U(z_1)+V(z_0)}}{e^{V(z_1)}} [y^t] \mathrm{COMPLEX}_t(z_0,\ z_1,\ y)\,.$$

Main ideas behind the asymptotics

Same tricks as for $G(n, M = \frac{1}{2}(n + \mu n^{2/3}))$ **[FKP'89,JKLP'93]**, random 2-XORSAT, bipartiteness **[Pittel, Yeum'10]**, ...

1. GF = Decompose as an infinite sum of **GFs.**
2. Bivariate saddle point method at the diagonal, i.e, $[{z_0}^n {z_1}^n]$.
3. Subcritical: first GF of the decomposition only contributes.
   "Bitrees" like components $\sim$ **Gauss**.
4. Critical: all **GFs** matter. **complex terms contribute as Airy summation in terms of** μ.

Example

Results in **[DPRRW'23]**

- Probability that $D(n, M = n + \mu n^{2/3})$ is acyclic. Probability that $D(n, M)$ is elementary.
- Probability to contain a SCC of given excess. Structures of SCC ...

Outline of the talk

- Introduction & motivations.
- Some tractable graph phase transitions.
- 2-SAT with generating functions.
- Conclusion and perspectives.

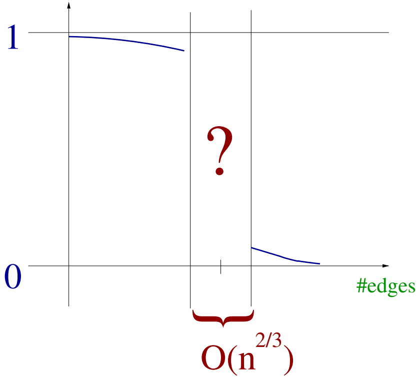
Random 2-SAT: what is inside the window?

- A long and rich problem: one of the K-SAT family with **proven threshold** **[Goerdt '92], [Chvátal, Reed '92], [de la Vega '92]**. Note that for $k > k_0$ (k_0 abs. constant), thresholds are given in **[Ding, Sly, Sun'22]**
- For 2-SAT, the best description of the window up to date **[Bollobàs, Borgs, Chayes, Kim, Wilson '01]**:
  the window of transition is of size $O(n^{2/3})$ i.e. let
  $p(n, m) = \mathbb{P}(\texttt{Formula}(n \texttt{ variables}, m \texttt{ clauses }) \texttt{ is SAT })$

$$p(n, m) = \begin{cases} \bullet\ 1 - O\left(\frac{n^2}{(n-m)^3}\right) \text{ as } n - m \gg n^{2/3} \\ \bullet\ O(1) \text{ as } |n - m| = O(n^{2/3}) \\ \bullet\ \exp\left(-\frac{(m-n)^3}{n^2}\right) \text{ as } m - n \gg n^{2/3} \end{cases}$$

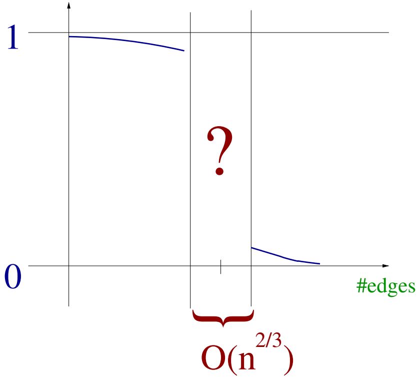
Random 2-SAT: what is known and what is next?

- Concretely **[Bollobàs et al. '01]**



Random 2-SAT: what is known and what is next?

- Concretely **[Bollobàs et al. '01]**



- A question of **[Flajolet '08]**. Note that physicists **[Deroulers, Monasson '06]** computed numerically (no inference)
  $\mathbb{P}\left[\text{formula}(10^6\text{variables},\ 10^6 \text{ clauses is SAT})\right] \sim 0.913\cdots$
- **Dovgal, de Panafieu, R.'23** : exact formulae (<u>**NO ASYMPTOTICS**</u>) encoded in combination of known generating functions.

Theorem. Let $S_{n,m}$ be the number of satisfiable 2-CNF with n Boolean literals and m clauses. Then,

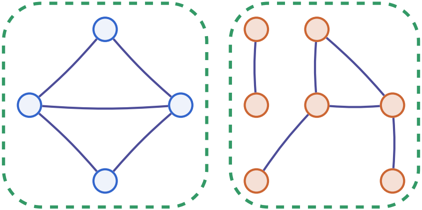
$$\ddot{S}(z,w) = \left[\sqrt{G(z,w) \odot_z \frac{1}{G(2z,w)}} \odot_z \ddot{Set}(z,w)\right] G\left(\frac{2z}{1+w}, w\right)$$

where

- $\ddot{S}(z,w) := \sum_{n=0}^{\infty} \sum_{m=0}^{2n(n-1)} S_{n,m} \frac{w^m}{(1+w)^{n^2}} \frac{z^n}{n!}$
- $\ddot{Set}(z,w) := \sum_{n=0}^{\infty} \frac{1}{(1+w)^{n^2}} \frac{z^n}{n!}$
- $G(z,w) := \sum_{n=0}^{\infty} (1+w)^{\binom{n}{2}} \frac{z^n}{n!}$ is the EGF of all simple graphs.
- $\odot_z$ is the exponential Hadamard product $\left(\sum_{n=0}^{\infty} a_n(\mathrm{w}) \frac{z^n}{n!}\right) \odot_z \left(\sum_{n=0}^{\infty} b_n(\mathrm{w}) \frac{z^n}{n!}\right) := \sum_{n=0}^{\infty} a_n(\mathrm{w}) b_n(\mathrm{w}) \frac{z^n}{n!}$.

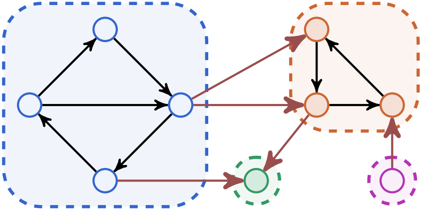
The cartesian product

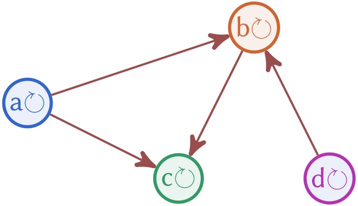
$$\left(a_0 + a_1\frac{z}{1!} + a_2\frac{z^2}{2!} + \dots\right)\left(b_0 + b_1\frac{z}{1!} + b_2\frac{z^2}{2!} + \dots\right) = c_0 + c_1\frac{z}{1!} + c_2\frac{z^2}{2!} + \dots$$



The convolution rule corresponding to EGF:

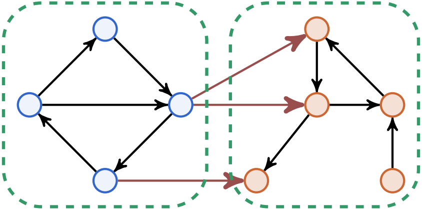
$$c_n = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}$$





- Components a⟳, b⟳, c⟳ and d⟳ are *strongly-connected* components.
- Components a⟳ and d⟳ are *source-like* components
- Component c⟳ is a *sink-like* component

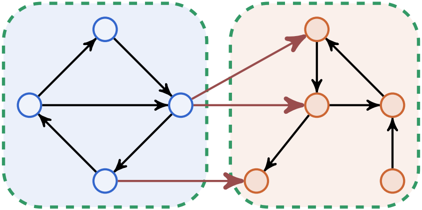
The arrow product



The arrow product convolution rule

Let $a_n = a_n(w)$, $b_n = b_n(w)$, $c_n = c_n(w)$,

$$\left(\sum_{n=0}^{\infty} a_n \frac{1}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!}\right)\left(\sum_{n=0}^{\infty} b_n \frac{1}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!}\right) = \sum_{n=0}^{\infty} c_n \frac{1}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!}$$



The convolution rule corresponding to Graphic GF:

$$c_n = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k} (1+w)^{k(n-k)}$$

Conversion/swapping between Exponential GF and Graphic GF

$$\widehat{A}(z,w) = \sum_{n=0}^{\infty} a_n(w) \frac{1}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!}, \quad A(z,w) = \sum_{n=0}^{\infty} a_n(w) \frac{z^n}{n!}$$

- **Exponential Hadamard product:**

$$\left(\sum_{n \geqslant 0} a_n(\mathrm{w}) \frac{z^n}{n!} \right) \odot_z \left(\sum_{n \geqslant 0} b_n(\mathrm{w}) \frac{z^n}{n!} \right) := \sum_{n \geqslant 0} a_n(\mathrm{w}) b_n(\mathrm{w}) \frac{z^n}{n!}$$

- **Exponential GF for graphs, Graphic GF for sets:**

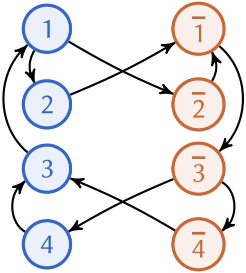
$$G(z,w) = \sum_{n \geqslant 0} (1+w)^{\binom{n}{2}} \frac{z^n}{n!}, \quad \widehat{Set}(z,w) = \sum_{n \geqslant 0} \frac{1}{(1+w)^{\binom{n}{2}}} \frac{z^n}{n!},$$

- **Conversion formulas:**

$$\boxed{A(z,w) = G(z,w) \odot_z \widehat{A}(z,w)} \quad \boxed{\widehat{A}(z,w) = \widehat{Set}(z,w) \odot_z A(z,w)}$$

Implication digraphs

$$\begin{cases} \neg x_1 \lor \neg x_2 & = 1 \\ \neg x_1 \lor x_2 & = 1 \\ x_3 \lor x_4 & = 1 \\ x_1 \lor \neg x_3 & = 1 \\ x_3 \lor \neg x_4 & = 1 \end{cases}$$



Replace each clause $x \lor y$ with two implications $\overline{x} \to y$ and $\overline{y} \to x$.

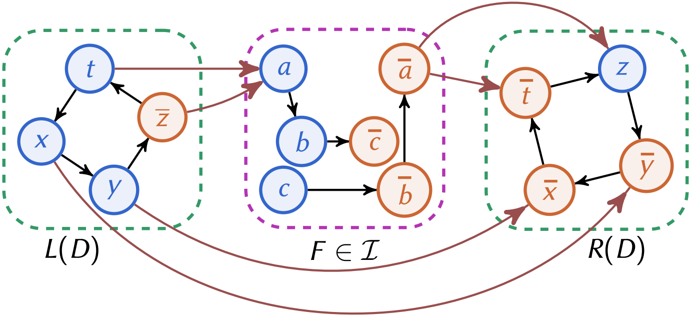
Proposition (folklore / [Aspvall, Plass, Tarjan '82])

2-CNF is satisfiable if and only if there is no *contradictory circuit.*

The above 2-CNF is not satisfiable

$$1 \longrightarrow \overline{2} \longrightarrow \overline{1} \longrightarrow \overline{3} \longrightarrow \overline{4} \longrightarrow 3 \longrightarrow 1$$

The implication product

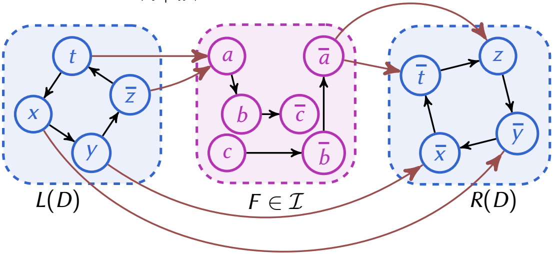


The implication product convolution rule

If $\widehat{A}$ is Graphic GF and $\ddot{B}$, $\ddot{C}$ are Implication GF then

$$\left(\sum_{n=0}^{\infty} a_n \frac{2^n}{(1+w)^{\binom{n+1}{2}}} \frac{z^n}{n!}\right)\left(\sum_{n=0}^{\infty} b_n \frac{1}{(1+w)^{n^2}} \frac{z^n}{n!}\right) = \sum_{n=0}^{\infty} c_n \frac{1}{(1+w)^{n^2}} \frac{z^n}{n!}$$

$$\widehat{A}\left(\tfrac{2z}{1+w}\right) \cdot \ddot{B}(z,w) = \ddot{C}(z,w)$$



Combinatorial convolution rule corresponding to Implication GF:

$$c_n = \sum_{k=0}^{n} \binom{n}{k} 2^k a_k b_{n-k} (1+w)^{k \cdot 2(n-k) + \binom{k}{2}}$$

Outline of the talk

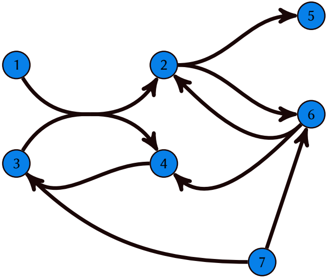
- Introduction & motivations.
- Some tractable graph phase transitions.
- 2-SAT with generating functions.
- Conclusion and perspectives.

Conclusion

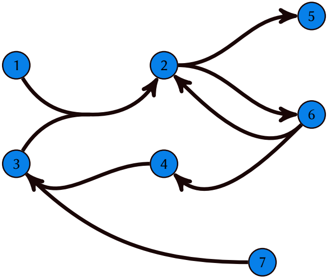
Enumerative approaches for similar problems

- From typical anatomy of the graphs, count only the main components (3-regular or cubic planar components)
- If we have some decompositions : enumerative/analytic combinatorics work well.
- **[de Panafieu, Ramos'16]** : new approaches on connected graphs with large excess.
- Random directed hypergraphs **(ongoing works)**.

Many Perspectives on Directed Random Hypergraphs



Nodes: $\{1, 2, 3, 4, 5, 6, 7\}$
Hyperarcs: $\{1, 3\} \rightarrow \{2, 4\}$, $\{2\} \rightarrow \{5, 6\}$, $\{6\} \rightarrow \{2, 4\}$, $\{7\} \rightarrow \{6\}$, $\{7, 4\} \rightarrow \{3\}$,
A cycle: $\{2, 6, 4, 3\}$
Introducing a **new kind of formal power series**,



A 3-uniform directed hypergraph with nodes $\{1, 2, 3, 4, 5, 6, 7\}$ and 4 hyperarcs $\{1, 3\} \rightarrow \{2\}$, $\{2\} \rightarrow \{5, 6\}$, $\{6\} \rightarrow \{2, 4\}$, $\{7, 4\} \rightarrow \{3\}$.

$$F(x, y) := \sum_{n=0}^{\infty} \frac{f_n(y)}{(1+y)^{\binom{n}{b}}} \frac{x^n}{n!},$$

we have **enumerative results** such as:

$$A(x, y) = \left(\sum_{n=0}^{\infty} \frac{(-x)^n}{n!\,(1+y)^{\binom{n}{b}}} \right)^{-1}$$

where $[x^n y^q]A(x, y)$ is related to the number of acyclic structures with n nodes and y arcs.

Stack of what to do?

Perspectives

Stack of what to do?

- Directed random hypergraphs (**`structural/analytic motivations/internships?`**).

- Mixtures of formula (**`algorithmic motivations`**):
  1. Asymptotics of 2-SAT.
  2. Mixture of $(2 - XOR, 2 - SAT)$.
     Interpolation between **coarse** and **sharp** phases transitions.
  3. $(2 + p)$-XORSAT where $p =$ 'proportion' of 3-XORSAT clauses). (Same kind of interpolation.)

Perspectives

Stack of what to do?

- Directed random hypergraphs (**`structural/analytic motivations/internships?`**).

- Mixtures of formula (**`algorithmic motivations`**):
  1. Asymptotics of 2-SAT.
  2. Mixture of $(2 - XOR, 2 - SAT)$.
     Interpolation between **coarse** and **sharp** phases transitions.
  3. $(2 + p)$-XORSAT where $p =$ 'proportion' of 3-XORSAT clauses). (Same kind of interpolation.)
  4. $(2 + p) - SAT$ is **extremely difficult.**
     Interpolation between **tractable/intractable** problems.