RESEARCH ANNOUNCEMENT: THE SUBCRITICAL CONSTANT OF 2-SAT IS $e^{-3/8e^2} \approx 0.06260588\ldots$*

SERGEY DOVGAL

ABSTRACT. Using a novel technique of **sum-representation** and inclusion-exclusion principle we obtain the constant C in the asymptotic expression for probability of satisfiability of 2-CNF

$$\mathbb{P}(F(n, n(1-\mu n^{-1/3})) \text{ is SAT}) = 1 - \frac{C}{\mu^3}, \quad n, \mu \to \infty$$

with explicit and unexpected value of $C = e^{-3/8e^2} \approx 0.0626\ldots$ which disproves Kim's "theorem" that $C = \frac{1}{16} = 0.0625$.

1. Introduction

We consider directed graphs on $2n$ vertices and m edges whose labels are partitioned into positive and negative:

$$V = \{1, 2, \ldots, n, \overline{1}, \overline{2}, \ldots, \overline{n}\}.$$

Any 2-CNF with n Boolean variables and m clauses can be represented in the form of an *implication digraph* where to every clause $(x \vee y)$ we assign two directed edges $(\overline{x} \to y)$ and $(\overline{y} \to x)$.

We say that a pair of edges in a digraph form a *conflict* if the first edge is of the form $x \to y$ and the second edge is of the form $\overline{y} \to \overline{x}$. We say that a digraph is *conflict-free* if there are no conflicting pairs. Every implication digraph can be represented as an edge union of two conflict-free digraphs in 2^m ways. We call every such representation *sum-representation*.

A formula is satisfiable if its implication digraph doesn't contain *contradictory cycles*, i.e. directed circuits containing both vertices x and $\overline{x}$. We count the contradictory circuits with the multiplicities of the corresponding *simplified paths* which will be explained below. The event that a formula is satisfiable is equivalent to the event that the number of contradictory circuits (with mentioned multiplicities) is equal to zero, which, by the inclusion-exclusion method, equal to

$$\mathbb{P}(\xi = 0) = 1 - \mathbb{E}\xi + \frac{1}{2!}\mathbb{E}\xi(\xi - 1) - \ldots$$

where ξ represents the number of contradictory circuits with the multiplicities of simplified paths.

We claim that in the *subcritical regime*, i.e. when $m = n(1 - \mu n^{-1/3})$, $n \to \infty$, $\mu \to \infty$ it holds

$$\mathbb{E}\xi = C\mu^{-3}, \quad \mathbb{E}\xi(\xi - 1) = O(\mu^{-6})$$

with the announced $C = e^{-3/8e^2}$. This immediately implies the announced result.

Date: April 1(!) 2019.

2. Contradictory component in the sum-representation

As shown in the paper "birth of the giant component", a random graph in its subcritical phase almost surely consists of trees and unicycles. We take a very particular sum-representation of a contradictory circuit containing a variable x and $\overline{x}$, namely a sequence of trees nested on a directed path

$$\mathcal{S} := x \rightsquigarrow \overline{x} \rightsquigarrow y \rightsquigarrow \overline{y}.$$

It is easy to see that this triple path, joined with its negated counterpart $y \rightsquigarrow \overline{y} \rightsquigarrow x \rightsquigarrow \overline{x}$, forms a contradictory circuit, and for every contradictory circuit we can detect such a sub-shape. Effectively, if we count formulas with such distinguished paths, we are counting with multiplicities.

Already, the method of exponential generating functions together with the saddle-point method introduced in [Janson, Knuth, Pittel, Łuczak], gives

$$\mathbb{P}(\text{a graphs contains } \mathcal{S}) \sim \frac{n!}{|F_{n,m}|}[z^n]\frac{U^{n-m}}{(n-m)!}e^{V(z)}\frac{1}{(1-T(z))^3}$$

where $T(z) = \sum_n \frac{z^n}{n!}n^{n-1}$ is the exponential generating function (EGF) for rooted trees, $U(z) = \sum_n \frac{z^n}{n!}n^{n-2}$ is the EGF for unrooted trees, and

$$V(z) = \frac{1}{2}\left[\log\frac{1}{1-T} - T - \frac{T^2}{2}\right]$$

is the EGF for unicyclic components (corresponding directed versions can be obtained by the substitution $z \mapsto 2z$). Coefficient extraction gives us $\Theta(\mu^{-3})$.

This needs a certain refinement: we didn't take into account that in the counted graphs the edge conflicts are forbidden. The last step is to account for this conflicts using inclusion-exclusion method and obtain the announced constant.

3. Marking edge conflicts

We are going to eliminate edge conflicts using inclusion-exclusion method. For this purpose, we add one more layer of inclusion-exclusion where ϑ represents the numberr of edge conflict pairs:

$$\mathbb{P}(\vartheta = 0) = 1 - \mathbb{E}\vartheta + \frac{1}{2!}\mathbb{E}\vartheta(\vartheta-1) - \ldots$$

It can be easily shown that edge conflicts happen most probably in the forest (in other parts the corresponding contribtion is negligible by a factor $n^{1/3}$). A tree with a marked conflict obtains a generating function

$$\iint (\partial_z U(z))^2 dz$$

instead of $U(z)$. After marking k pairs we need to account for the following adjustments: $3\binom{n}{2}$ is the total possible number of non-double pairs of edges. We approximate it by just $\frac{3}{2}n^2$. After marking two trees we adjust by a factor $n^2 \left.\frac{T(z)^2}{U(z)^2z^2}\right|_{z=e^{-1}}$ where substitution yields $T = 1$, $z = e^{-1}$, $U = 1/2$. We also multiply by $(2n)^{-4}$ because of the double (quadruple) factorial in the denominator.

Multiplying over the total contribution leads to massive cancellations and

$$term_k \sim \frac{1}{k!}\left(\frac{3}{8}e^2\right)^k, \quad \sum_k (-1)^k term_k \sim e^{-3e^2/8} \sim 0.0626058824$$

Details will follow elsewhere.

E-mail address: `dovgal-at-lipn.fr`