The Impact of Heterogeneity and Geometry on the Proof Complexity of Random Satisfiability*

Thomas Bläsius† Tobias Friedrich‡ Andreas Göbel‡ Jordi Levy§ Ralf Rothenberger‡

Abstract

Satisfiability is considered the canonical NP-complete problem and is used as a starting point for hardness reductions in theory, while in practice heuristic SAT solving algorithms can solve large-scale industrial SAT instances very efficiently. This disparity between theory and practice is believed to be a result of inherent properties of industrial SAT instances that make them tractable. Two characteristic properties seem to be prevalent in the majority of real-world SAT instances, heterogeneous degree distribution and locality. To understand the impact of these two properties on SAT, we study the proof complexity of random k-SAT models that allow to control heterogeneity and locality. Our findings show that heterogeneity alone does not make SAT easy as heterogeneous random k-SAT instances have superpolynomial resolution size. This implies intractability of these instances for modern SAT-solvers. On the other hand, modeling locality with an underlying geometry leads to small unsatisfiable subformulas, which can be found within polynomial time.

A key ingredient for the result on geometric random k-SAT can be found in the complexity of higher-order Voronoi diagrams. As an additional technical contribution, we show an upper bound on the number of non-empty Voronoi regions, that holds for points with random positions in a very general setting. In particular, it covers arbitrary $\mathfrak{p}$-norms, higher dimensions, and weights affecting the area of influence of each point multiplicatively. Our bound is linear in the total weight. This is in stark contrast to quadratic lower bounds for the worst case.

1 Introduction

Propositional satisfiability (SAT) is arguably among the most-studied problems for both theoretical and practical research. Nonetheless, the gap between theory and practice is huge. In theory, SAT is the prototypical hard problem and hardness of other problems is shown via reductions from SAT. Achieving even a running time of $O(2^{cn})$ for any $c < 1$ and n variables would be a major breakthrough and a somewhat surprising one at that. On the contrary, reductions to SAT are used to solve various problems appearing in practice, as state-of-the-art SAT solvers can easily handle industrial instances with millions of variables.

This theory–practice gap does not come from the lack of a sufficiently precise theoretical analysis of modern SAT solvers. They are actually provably slow on most instances, i.e., drawing an instance uniformly at random yields a hard instance with probability tending to 1 for $n \to \infty$, if the clause-variable ratio is not too low or way too high [7, 20]. Instead, the discrepancy comes from the fact that industrial instances have properties that make them easier than worst-case instances. In 2014, Vardi [52] wrote that "we have no understanding of why the specific sets of heuristics employed by modern SAT solvers are so effective in practice" and that we need this understanding to successfully advance SAT solving further.

In recent years, scientists have been studying properties of industrial SAT instances to gain this understanding. By modeling SAT instances as graphs, e.g., with edges indicating inclusion of variables in clauses, one can benefit from the extensive research conducted in the field of network science. Two properties commonly observed in real-world networks are heterogeneity and locality. *Heterogeneity* refers to the degree distribution, meaning that vertices have strongly varying degrees. In fact, one usually observes a heavy-tailed distribution with many vertices of low degree and few vertices of high degree. A common assumption is a power-law distribution [53], where the number of vertices of degree k is roughly proportional to $k^{-\beta}$. The constant β is called the *power-law exponent*. *Locality* refers to the fact that edges tend to connect vertices that are close in the sense that they remain well connected even when ignoring their direct connection. This can also be seen as having strong community structures, with high connectivity within communities and loose ties between communities.

With respect to these two properties, industrial SAT instances are similar to real-world networks. In many cases, the variable frequencies are heterogeneous [1]

*This work is partially funded by the project *Scale-Free Satisfiability* (project no. 416061626) of the German Research Foundation (DFG).

†Karlsruhe Institute of Technology, Germany
‡Hasso Plattner Institute, University of Potsdam, Germany
§IIIA, CSIC, Campus UAB, 08193 Bellaterra, Spain

and there is a high level of locality [2]. The latter is often measured in terms of modularity. Inspired by network science, researchers have studied models that resemble industrial instances with respect to these properties. Particularly, Ansótegui et al. [3] introduced a power-law SAT model for heterogeneous instances, which has been theoretically studied in terms of satisfiability thresholds [28–30]. Moreover, Giráldez-Cru and Levy [35] introduced a model in which variable weights lead to heterogeneity while an underlying geometry facilitates locality. Comparing this to network models, the former model [3] is the SAT-variant of Chung-Lu graphs [18, 19]. The latter [35] is based on the popularity-similarity model [47], which is closely related to hyperbolic random graphs [39] and geometric inhomogeneous random graphs [17].

Besides serving as somewhat realistic benchmarks for SAT competitions [34], these SAT models can be used to study solver behavior depending on heterogeneity and locality. One can experimentally observe that a high level of heterogeneity improves the performance of SAT solvers that also perform well on industrial instances [3, 11]. Moreover, locality seems very beneficial as solvers appear to implicitly use the locality of a given instance [35]. This coincides with the findings of experiments on actual industrial instances that show that the locality (measured using modularity) of an instance is a good predictor for solver performance [46, 54, 55].

Up to date, there are no theoretical results supporting these experimental observations. On the contrary, it has been shown that instances generated by the community attachment model [33], which enforces a community structure, are hard for modern SAT solvers [43]. With this paper, we provide a theoretical foundation that matches the observations in practice by studying the proof complexity of k-SAT instances (for constant k) drawn from the power-law SAT model, and from a very general model with underlying geometry. The former was introduced by Ansótegui et al. [3], the latter is a generalization of geometric model by Giráldez-Cru and Levy [35] in the same way as geometric inhomogeneous random graphs [17] are a generalization of hyperbolic random graphs [39]. Our findings are that heterogeneous instances are hard asymptotically almost surely[1] in that modern SAT solvers require superpolynomial or even exponential running time to refute unsatisfiable instances. On the contrary, instances with a high level of locality facilitated by an underlying geometry are a. a. s. easy to solve. Our results focus on unsatisfiable instances, i.e., on the case where a solver has to prove that no satisfying assignment exists. This is typically much harder than finding a satisfying assignment, making the unsatisfiable regime arguably more relevant. Besides these results on SAT, we provide insights on the complexity of weighted higher-order Voronoi diagrams in higher dimensions, which is of independent interest.

The power-law and geometric models both mimic specific properties observed in industrial instances while trying to make as little additional assumptions as possible. Though this makes the resulting instances arguably more realistic than, e.g., instances drawn uniformly at random, we want to stress that even the geometric model is far from a perfect representation of industrial instances. Thus, our results do not claim to completely explain the efficiency of modern SAT solvers on industrial instances. However, to the best of our knowledge, we provide the first theoretical result that links a high level of locality to provably more tractable instances, which we believe to be a first step towards closing the theory–practice gap.

Outline. We state and discuss our main results and technical contributions in Section 2. Formal definitions are in Section 3. A short outline of our core arguments is in Section 4. The detailed proofs can be found in the full version [12].

2 Results, Technical Contribution, Discussion

In this section, we state our results and discuss the contribution, also in context to previous results. To make the results understandable, we briefly discuss, e.g., the probability distributions over SAT formulas we study. These are short and not meant to be formal definitions. For complete definitions, see Section 3.

2.1 Power-Law SAT. The power-law SAT model has four parameters: the number of variables n, the number of clauses m, the number k of variables appearing in each clause, and a power-law exponent β. To draw a formula, power-law weights with exponent β are assigned to the variables and then each clause is generated independently by drawing k variables without repetition using probabilities proportional to the weights. Each literal is negated with probability $1/2$.

To discuss our first main contribution, let Φ be a formula drawn from the power-law model with density above the satisfiability threshold, i.e., Φ is a. a. s. unsatisfiable. We show that, although it is highly likely that Φ is unsatisfiable, it is highly unlikely that modern SAT solvers can figure that out in polynomial time. We prove this via a lower bound for the *resolution width*, which is the size of the largest clause appearing in a resolution proof. A lower bound w for the width implies lower

[1] *Asymptotically almost surely (a. a. s.)* refers to a probability that tends to 1 for $n \to \infty$. *With high probability (w. h. p.)* refers to the stronger requirement that the probability is in $1 - O(1/n)$.

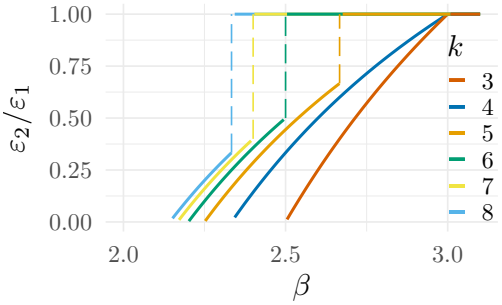


Figure 1: Exponent of the bound (i) in Theorem 2.1. Dashed vertical lines show where the bound (iv) takes over.

bounds on the size of resolution proofs [7]: every resolution proof has size $\exp(\Omega(w^2/n))$ and every tree-like resolution proof has size $\exp(\Omega(w))$. The lower bound on the resolution size directly translates to a matching lower bound on the running time of conflict-driven clause learning (CDCL) solvers [6, 48]. For Davis–Putnam–Logemann–Loveland (DPLL) solvers, which use tree-like resolution, the bounds are even stronger. We note that our bound does not only hold above the satisfiability threshold, where Φ is a. a. s. unsatisfiable, but also at the threshold, where it is unsatisfiable with constant probability.

THEOREM 2.1. *Let Φ be a random power-law k-SAT formula with n variables, $m \in \Omega(n)$ clauses, $k \geq 3$, and power-law exponent $\beta > \frac{2k-1}{k-1}$. Let $\Delta = m/n$ be large enough so that Φ is unsatisfiable at least with constant probability. Let ε, $\varepsilon_1, \ldots, \varepsilon_3$ be constants with $\varepsilon > 0$, $\varepsilon_1 = \frac{k-\varepsilon}{2} - 1 > 0$, $\varepsilon_2 = (k-\varepsilon)\frac{\beta-2}{\beta-1} - 1 > 0$, and $0 < \varepsilon_3 < (\frac{k}{2} - 1)\frac{\beta-2}{\beta-1} - 1$. For the resolution width w of Φ, it holds a. a. s. that:*

(i) If $\beta \in \left(\frac{2k-1}{k-1}, 3\right)$ and $\Delta \in o\left(n^{\varepsilon_2}\right)$, then $w \in \Omega\left(n^{\varepsilon_2/\varepsilon_1}\Delta^{-1/\varepsilon_1}\right)$.

(ii) If $\beta = 3$ and $\Delta \in o\left(n^{\varepsilon_1}/\log^{1+\varepsilon_1} n\right)$, then $w \in \Omega\left(n \cdot \Delta^{-1/\varepsilon_1}/\log^{1+1/\varepsilon_1} n\right)$.

(iii) If $\beta > 3$ and $\Delta \in o\left(n^{\varepsilon_1}\right)$, then $w \in \Omega\left(n \cdot \Delta^{-1/\varepsilon_1}\right)$.

(iv) If $\beta > \frac{2k-2}{k-2}$ and $\Delta \in o\left(n^{\varepsilon_3}/\log^{\varepsilon_3} n\right)$, then $w \in \Omega\left(n \cdot \Delta^{-1/\varepsilon_3}\right)$.

The above lower bounds allow the density Δ to be super-constant (even polynomially), which is asymptotically above the satisfiability threshold. For the sake of simplicity, assume Δ to be constant in the following. Starting at the bottom (iii, iv), we get a linear bound for w if β is sufficiently large, i.e., greater than 3 or $(2k-2)/(k-2)$. For $\beta = 3$ (ii), the bound is still almost linear. Note that these results in particular imply exponential lower bounds on the resolution size and thus on the running time of CDCL and DPLL. For smaller β (i), we get a polynomial bound for the width with exponent $\varepsilon_2/\varepsilon_1$; see Figure 1 for a plot with ε close to 0.

Interestingly enough, our bounds only hold for power law exponents $\beta > \frac{2k-1}{k-1}$. This is complemented by a previous result [29], which shows that the satisfiability threshold of power-law random k-SAT is at density $\Delta = \Theta(1)$ for power law exponents $\beta > \frac{2k-1}{k-1}$ and that asymptotically almost surely instances with constant constraint densities are trivially unsatisfiable for power law exponents $\beta < \frac{2k-1}{k-1}$. Thus, the resolution width is constant in the latter case.

Part iv of Theorem 2.1 is derived via lower bounds on the bipartite expansion of the clause-variable incidence graph of these instances. These results can be of independent interest for hypergraphs with edge size k and for random $(0,1)$-matrices. Additionally, these expansion properties yield lower bounds for the clause space complexity, which in turn gives lower bounds on the tree-like resolution of such formulas. More precisely, this gives an exponential lower bound for the tree-like resolution for $\beta > \frac{2k-3}{k-2}$. This is an improvement of the bound obtained via the resolution width.

It is interesting to note that this result on the non-geometric model supports the claim that locality is a crucial factor for easy SAT instances. The lower bounds for the power-law model are solely based on the fact that every set of clauses covers a comparatively large set of variables. In other words, we only use that there are no clusters of clauses with similar variables, i.e., we explicitly use the lack of locality.

2.2 Geometric SAT. The geometric model has the following parameters: n, m, and k have the same meaning as for the power-law model. Moreover, w is a weight function assigning each variable v a weight w_v and T is the so-called temperature that controls the strength of locality by varying the impact of the geometry. As underlying geometric space, we use the d-dimensional torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ (see Section 3) equipped with a $\mathfrak{p}$-norm with $\mathfrak{p} \in \mathbb{N}^+ \cup \infty$. To draw a formula, the variables and clauses are assigned random positions in $\mathbb{T}^d$. Then, for each clause, k variables are drawn without repetition with probabilities depending on the variable weight and on the geometric distance between clause and variable. In the extreme case of $T = 0$, each clause deterministically includes the k closest variables (where

closeness is a combination of geometric distance and weight), while increasing the temperature T increases the probability for the inclusion of more distant variables. For $T \to \infty$, the model converges to uniform random SAT. Note that the weights are a parameter of the model and not drawn randomly. We have the following theorem, where W denotes the sum of all variable weights. The condition on the weights is in particular satisfied by power-law distributed weights.

THEOREM 2.2. *Let Φ be a formula with n variables and $m \in \Theta(n)$ clauses drawn from the weighted geometric model with ground space $\mathbb{T}^d$ equipped with a $\mathfrak{p}$-norm, temperature $T < 1$, $W \in O(n)$, and $w_v \in O(n^{1-\varepsilon})$ for every $v \in V$ and any constant $\varepsilon > 0$. Then, Φ contains a. a. s. an unsatisfiable subformula of constant size, which can be found in $O(n \log n)$ time.*

To briefly explain how we prove this, consider a simplified version where variables and clauses are points in the Euclidean plane and each clause contains the k variables geometrically closest to it (temperature $T = 0$). Now consider the equivalence relation obtained by defining two points of the plane equivalent if and only if they have the same set of k closest variables. The equivalence classes of this relation are the regions of the order-k Voronoi diagram of the variable positions. With this connection, we can use upper bounds on the complexity of order-k Voronoi diagrams [41] to prove the existence of small and easy to find unsatisfiable subformulas. We note that this result is of asymptotic nature. In particular for small densities, the number of variables n has to be very large before the instances actually get as easy as stated in Theorem 2.2. Nevertheless, this results strongly suggests that an underlying geometry makes SAT instances more tractable.

To extend the above argument to the general statement in Theorem 2.2, we extend the complexity bounds for order-k Voronoi diagrams in various ways; see next section for more details. Moreover, for non-zero temperatures, clauses no longer include exactly the k closest variables but can, in principle, consist of any set of k variables. However, we can show that, with high probability, a linear fraction of clauses behaves as in the $T = 0$ case. We note that analyses of similar structures, such as hyperbolic random graphs, are often restricted to the simpler but less realistic $T = 0$ case, e.g., [9, 10, 13, 45]. We believe that our analysis provides insights on the non-zero temperature case that can be helpful for such related questions.

We note that our results seem to contradict the results of Mull et al. [43], stating that (i) a strong community structure is not sufficient to have tractable SAT instances and that (ii) the community attachment model [33], which enforces a community structure, generates hard instances. However, at a closer look, this is not a contradiction at all. Though measuring the community structure, e.g., via modularity, is a good indicator for locality, the concept of locality goes deeper. If the instance can be partitioned such that there are strong ties within each partition and loose ties between partitions, then the instance has a strong community structure. However, to have a high level of locality, this concept has to hierarchically repeat on different levels of magnitude, i.e., there needs to be community structure within each partition and between the partitions. To state this slightly differently, consider locality based on a notion of similarity between objects (here: variables or clauses). In this paper, we use distances between random points in a geometric space as a measure for similarity, which gives us a continuous range of more or less similar objects. In contrast to that, in the above mentioned papers focusing on a flat community structure [33, 43], similarity is a binary equivalence relation: two objects are either similar or they are not.

2.3 Voronoi Diagrams. Consider a finite set of points, called sites, in a geometric space. The most commonly studied type of Voronoi diagram assumes the 2-dimensional Euclidean plane as ground space and has one Voronoi region for each site, containing all points closer to this site than to any other site. We deviate from this default setting in four ways: (i) We allow an arbitrary constant dimension d, where the ground space is the torus or a hypercube in $\mathbb{R}^d$. (ii) We consider the order-k Voronoi diagram, which has for every subset A of sites with $|A| = k$ a (possibly empty) Voronoi region containing all points for which A are the k nearest sites. The number of non-empty order-k Voronoi regions is called the *complexity* of the diagram. (iii) The sites have multiplicative weights that scale the influence of the different sites. Without loss of generality, we assume the weights to be scaled such that the minimum is 1. (iv) We allow the $\mathfrak{p}$-norm for arbitrary $\mathfrak{p} \in \mathbb{N}^+ \cup \infty$.

THEOREM 2.3. *Let S be a set of n sites with minimum weight 1, total weight W, and random positions on the d-dimensional torus equipped with a $\mathfrak{p}$-norm, for constant d. For every fixed k, the expected number of regions of the weighted order-k Voronoi diagram of S is in $O(W)$. The same holds for random sites in a hypercube.*

To set this result into context, we briefly discuss previous work on the complexity of Voronoi diagrams in different settings. See the book by Aurenhammer et al. [5] for a general overview on Voronoi diagrams. To this end, we use the following theorem that relates the complexity in terms of Voronoi regions (which is what

we are concerned with in this paper) with the complexity in terms of vertices.[2]

THEOREM 2.4. *Let S be a set of n weighted sites in general position in $\mathbb{R}^d$ equipped with a $\mathfrak{p}$-norm. If the order-k Voronoi diagram has ℓ vertices, then the order-$(k+d)$ Voronoi diagram has $\Omega(\ell)$ non-empty regions.*

We note that, using insights from previous work, this theorem is not hard to prove. One basically has to generalize the result by Lê [40] bounding the number of d-spheres going through $d+1$ points in d-dimensional space to weighted sites, and then observe how the Voronoi diagram changes in the construction by Lee [41] for $d = 2$, when going from order-k to order-$(k+1)$. However, we are not aware of previous work stating this connection between vertices and non-empty regions in higher orders explicitly.

The four above-mentioned generalizations of the basic Voronoi diagram (higher dimension, higher order, multiplicative weights, and different $\mathfrak{p}$-norms) have all been considered before. However, to the best of our knowledge, not all of them together.

Higher-order Voronoi diagrams have been introduced by Shamos and Hoey [51]. Lee [41] showed that the order-k Voronoi diagram in the plane (unweighted with Euclidean metric) has complexity $O(k\,(n-k))$ (in terms of number of regions), which is linear for constant k. For the 1- and ∞-norm, Liu et al. [42] improved this bound to $O(\min\{k\,(n-k), (n-k)^2\})$. Closely related to the 1-norm, Gemsa et al. [32] showed similar complexity bounds for higher-order Voronoi diagrams on transportation networks of axis-parallel line segments. Bohler et al. [14] show an upper bound of $2k\,(n-k)$ for the much more general setting of abstract Voronoi diagrams. There, the metric is replaced by curves separating pairs of sites such that certain natural (but rather technical) conditions are satisfied. One obtains normal Voronoi diagrams when using perpendicular bisectors for these curves. This in particular shows that the $2k\,(n-k)$ bound on the number of regions in the order-k Voronoi diagram holds for arbitrary $\mathfrak{p}$-norms in 2-dimensional space and for the hyperbolic plane. As the hyperbolic plane is closely related to 1-dimensional space with sites having multiplicative power-law weights [17], we suspect that the bound by Bohler et al. [14] also covers this case.

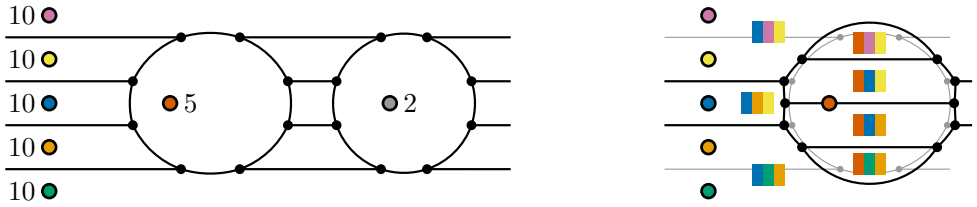
In general one can say that higher-order Voronoi diagrams of unweighted sites in 2-dimensional space are well-behaved in that they have linear complexity. This still holds true for arbitrary $\mathfrak{p}$-norms. However, this picture changes for weighted sites or higher dimensions.

Voronoi diagrams with multiplicative weights were first considered by Boots [16][3] due to applications in economics. Beyond that, multiplicatively weighted Voronoi diagrams have applications in sensor networks [21], logistics [31] and the growth of crystals [22]. However, even in the most basic setting of 2-dimensional Euclidean space and order 1, weighted Voronoi diagrams can have quadratic complexity [4] (in terms of number of vertices). This comes from the fact that Voronoi cells are not necessarily connected; see Figure 2a for the construction of Aurenhammer and Edelsbrunner [4] that proves the lower bound. With Theorem 2.4, and as illustrated in Figure 2, this implies that even the order-3 Voronoi diagram of weighted sites in 2-dimensional Euclidean space has a quadratic number of non-empty regions. As a special case, Theorem 2.3 shows that this complexity is only linear in the total weight for sites positioned randomly in the unit square. Moreover, this also implies that the number of vertices of the corresponding order-1 Voronoi diagram is linear. This nicely complements the result by Har-Peled and Raichel [37], who show that the expected complexity of order-1 Voronoi diagrams of sites in 2-dimensional Euclidean space with random weights is $O(n \operatorname{polylog} n)$. Only recently, Fan and Raichel [27] showed that sites with weights chosen randomly form a constant-sized set of possible weights yield Voronoi diagrams with linear complexity. Moreover, more closely related, they show that the Voronoi diagram of sites with arbitrary weights and with random positions chosen in the unit square has linear complexity in expectation. We are not aware of any results concerning the complexity of Voronoi diagrams when combining multiplicative weights with higher dimension, higher order or other norms.

For higher dimensions, even normal (first order, unweighted) Voronoi diagrams in 3-dimensional Euclidean space can have $\Theta(n^2)$ [38, 50] vertices. Theorem 2.4 thus implies that the order-4 Voronoi diagram has a quadratic number of non-empty regions. Moreover, the complexity of higher-order Voronoi diagrams in higher dimensions has been considered before by Mulmuley [44], who obtains polynomial bounds with the degree of the polynomial depending on the dimension. Our Theorem 2.3 in particular shows that this complexity is much lower, namely linear, for the hypercube with randomly positioned sites. Moreover, via Theorem 2.4 this gives a linear bound on number of vertices in the normal

[2] Although the Voronoi regions are not necessarily polytopes in the weighted setting, we adopt the notion for polytopes and call the corners of Voronoi regions *vertices*. I.e., vertices are the 0-dimensional elements (a.k.a. points) of the boundary, where higher-dimensional elements (a.k.a. edges, faces, etc.) intersect. They are represented as small black dots in Figure 2.

[3] In this paper, Voronoi regions are called *Thiessen polygons*.



(a) **(b)**

Figure 2: **(a)** Weighted Voronoi diagram (order-1) of the colored sites. Continuing the construction with $n/2$ high-weight sites on the left and $n/2$ low-weight sites towards the right yields $\Theta(n^2)$ vertices (small black dots). Note that each vertex lies on the boundary of three regions and has thus equal weighted distance to its three closest sites. **(b)** The order-3 Voronoi diagram for the same sites (excluding one). The colored boxes indicate the three closet sites. The order-1 diagram is shown in the background. Each order-1 vertex lies in the interior of an order-3 region as it has equal weighted distance to its three closest sites. As at most two order-1 vertices share an order-3 region, we get $\Omega(n^2)$ order-3 regions. Theorem 2.4 generalizes this observation.

order-1 Voronoi diagram in higher dimensions. We note that this special case of our result coincides with a previous result by Bienkowski et al. [8]. Similarly, Dwyer [24] showed that sites drawn uniformly from a higher dimensional unit sphere (instead of a hypercube) yield Voronoi diagrams of linear complexity in expectation. Moreover, due to Golin and Na [36] and Driemel et al. [23], the same is true for random sites on 3-dimensional polytopes and random sites on polyhedral terrains, respectively. Thus, though higher dimensional Voronoi diagrams can be rather complex in the worst case, these results indicate that one can expect most instances to be rather well behaved. An alternative explanation of why the complexity of practical instances is lower than the worst-case indicates is given by Erickson [25, 26], who studies the complexity of 3-dimensional Voronoi diagrams depending on the so-called spread of the sites.

The above results for higher dimensional Voronoi diagrams consider the Euclidean norm. For general $\mathfrak{p}$-norms, Lê [40] showed that the complexity of the Voronoi diagram is bounded by $O(n^c)$, where c is a constant independent of $\mathfrak{p}$ but dependent on the dimension d. With the same argument as above, Theorem 2.3 together with Theorem 2.4 implies a linear bound for this complexity that holds in expectation. Moreover, Boissonnat et al. [15] show more precise bounds of $\Theta(n^{\lceil d/2 \rceil})$ and $\Theta(n^2)$ for the ∞- and the 1-norm, respectively. Again, our result implies linear bounds for random sites in this setting.

3 Formal Definitions

Here we provide formal definitions for all concepts we use throughout the paper, including the power-law and geometric random SAT models, and Voronoi diagrams.

Power-Law Random k-SAT. The power-law model can be defined via the more general *non-uniform model*. To draw a k-SAT formula from the non-uniform model, let n and m be the number of variables and clauses, respectively, and let $w_1, \dots, w_n$ be variable weights. We sample m clauses independently at random. Each clause is sampled by drawing k variables without repetition with probabilities proportional to their weights. Then each of the k variables is negated independently at random with probability $1/2$.

The *power-law model* for a power-law exponent $\beta > 2$ is an instantiation of the non-uniform model with discrete power-law weights

$$w_i = i^{-\frac{1}{\beta-1}}.$$

Graph Representation and Expansion. Let Φ be a SAT-formula with variable set V and clause set C. The *clause-variable incidence graph* $G(\Phi)$ of Φ has vertex set $C \cup V$, with an edge between a clause and a variable if and only if the clause contains the variable. Clearly, $G(\Phi)$ is bipartite. It is an (r, c)-*bipartite expander* if for all $C' \subset C$ with $|C'| \leq r$ it holds that $|N(C')| \geq (1+c)|C'|$, where $N(C')$ is the neighborhood of C'.

Geometric Ground Space. We regularly deal with points with random positions in some geometric space. With *random point*, we refer to the uniform distribution in the sense that the probability for a point to lie in a region A is proportional to its volume $\mathrm{vol}(A)$. For this to work, the volume of the ground space has to be bounded. Canonical options are, e.g., a unit-hypercube or a unit-ball. These, however, lead to the necessity of special treatment for points close to the boundary, which makes the analysis more tedious without giving additional insights. To circumvent this, we use a torus as ground space, which is completely symmetric.

The *d-dimensional torus* $\mathbb{T}^d$ is defined as the d-dimensional hypercube $[0,1]^d$ in which opposite borders are identified, i.e., a coordinate of 0 is identical to a coordinate of 1.[4] It is equipped with the $\mathfrak{p}$-norm as metric, for arbitrary but fixed $\mathfrak{p} \in \mathbb{N}^+ \cup \{\infty\}$. To define it formally for the torus, let $\boldsymbol{p} = (p_1, \dots, p_d)$ and $\boldsymbol{q} = (q_1, \dots, q_d)$ be two points in $\mathbb{T}^d$. The circular difference between the ith coordinates is $|p_i - q_i|_\circ = \min\{|p_i - q_i|, 1 - |p_i - q_i|\}$. With this, the *distance* between $\boldsymbol{p}$ and $\boldsymbol{q}$ is

$$\|\boldsymbol{p} - \boldsymbol{q}\| = \begin{cases} \sqrt[\mathfrak{p}]{\sum_{i\in[d]} |p_i - q_i|_\circ^{\mathfrak{p}}} & \text{for } \mathfrak{p} \neq \infty, \\ \max_{i\in[d]}\{|p_i - q_i|_\circ\} & \text{for } \mathfrak{p} = \infty. \end{cases}$$

Random Points. We obtain the uniform distribution for a point $\boldsymbol{p} = (p_1, \dots, p_d)$ by drawing each coordinate p_i uniformly at random from $[0,1]$. For two random points $\boldsymbol{p}$ and $\boldsymbol{q}$, their distance $\|\boldsymbol{p} - \boldsymbol{q}\|$ is a random variable. Let $F_{\text{dist}}(x)$ be its *cumulative distribution function (CDF)*, i.e., $F_{\text{dist}}(x) = \Pr\left[\|\boldsymbol{p} - \boldsymbol{q}\| \leq x\right]$. To determine $F_{\text{dist}}(x)$, fix the position of $\boldsymbol{p}$. Then, for $x \leq 0.5$, the set of points of distance at most x to $\boldsymbol{p}$ is simply the ball $B_{\boldsymbol{p}}(x)$ of radius x around $\boldsymbol{p}$, yielding

$$\begin{aligned} F_{\text{dist}}(x) &= \text{vol}(B_{\boldsymbol{p}}(x)) \\ &= \Pi_{d,\mathfrak{p}} \cdot x^d \quad \text{for } 0 \leq x \leq 0.5, \\ \text{with } \Pi_{d,\mathfrak{p}} &= \frac{\left(2\Gamma\left(1/\mathfrak{p}+1\right)\right)^d}{\Gamma\left(d/\mathfrak{p}+1\right)}. \end{aligned} \tag{3.1}$$

where Γ is the gamma function. Note that $\Pi_{d,\mathfrak{p}}$ only depends on d and $\mathfrak{p}$ but is constant in x. Moreover $\Pi_{2,2} = \pi$ (thus the name Π), and $\Pi_{d,\infty} = \lim_{\mathfrak{p}\to\infty} \Pi_{d,\mathfrak{p}} = 2^d$. For distances $x > 0.5$, the formula for $F_{\text{dist}}(x)$ is more complicated (we basically have to subtract the parts reaching out of the hypercube). However, for our purposes, it suffices to know $F_{\text{dist}}(x)$ for $x \leq 0.5$ and use the obvious bound $F_{\text{dist}}(x) \leq 1$ for $x > 0.5$.

Weighted Points and Distances. We regularly consider a fixed set of n points equipped with weights, which we call *sites*. For a site $\boldsymbol{s}_i$ with weight w_i, the *weighted distance* of a point $\boldsymbol{p}$ to $\boldsymbol{s}_i$ is $\|\boldsymbol{s}_i - \boldsymbol{p}\|/w_i^{1/d}$. For a fixed value x, the set of points with weighted distance at most x are the points with $\|\boldsymbol{s}_i - \boldsymbol{p}\| \leq x w_i^{1/d}$. Note that the volume of this set is proportional to w_i. Intuitively, the region of influence of a site is thus proportional to its weight. To simplify notation in some places, we define *normalized weights* $\omega_i = w_i^{1/d}$.[5]

Geometric Random k-SAT. In the *geometric model*, we sample positions for the variables and clauses uniformly at random in the d-dimensional torus $\mathbb{T}^d$. For $v \in V$ and $c \in C$, we use $\boldsymbol{v}$ and $\boldsymbol{c}$ to denote their positions, respectively. Let $w_1, \dots, w_n$ be *variable weights* that are normalized such that the smallest weight is 1. Moreover, let $W = \sum_{v=1}^n w_v$. For a clause c and a variable v, define the *connection weight*

$$X(c,v) = \left(\frac{w_v}{\|\boldsymbol{c} - \boldsymbol{v}\|^d}\right)^{1/T}.$$

This is the reciprocal of the weighted distance between $\boldsymbol{v}$ and $\boldsymbol{c}$ raised to the power d/T. The k variables for the clause c are drawn without repetition with probabilities proportional to $X(c,v)$. Among all possible combinations, we choose which of the k variables to negate uniformly at random, without repetition if possible, i.e., we only get the same clause twice if we have more than 2^k clauses with the same variable set. For $T \to 0$ the model converges to the threshold case where c contains the k variables with smallest weighted distance.

The connection weight $X(c,v)$ is a random variable. We denote the CDF of $X(c,v)$ with $F_X(x)$. With the CDF for the distance between two random points in Equation (3.1), we obtain the following [12]:

$$F_X(x) = 1 - \Pi_{d,\mathfrak{p}} w_v x^{-T} \quad \text{for } x \geq \left(2^d w_v\right)^{1/T}. \tag{3.2}$$

Voronoi Diagrams. Let $S = \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_n\}$ be a set of sites with weights $w_1, \dots, w_n$. A point $\boldsymbol{p}$ belongs to the *(open) Voronoi region* of a site $\boldsymbol{s}_i$ if its weighted distance to $\boldsymbol{s}_i$ is smaller than its weighted distance to any other site. The collection of all Voronoi regions is the *Voronoi diagram* of S. *Order-k Voronoi regions* are defined analogously for subsets $A \subseteq S$ with $|A| = k$, i.e., the region of A contains a point $\boldsymbol{p}$ if and only if the weighted distances of $\boldsymbol{p}$ to all sites in A is smaller than the weighted distance to any site not in A. More formally, $\boldsymbol{p}$ belongs to the order-k Voronoi region of A if there exists a radius r such that $\|\boldsymbol{s}_i - \boldsymbol{p}\| \leq \omega_i r$ for $\boldsymbol{s}_i \in A$ and $\|\boldsymbol{s}_i - \boldsymbol{p}\| > \omega_i r$ for $\boldsymbol{s}_i \notin A$. Note that the order-$k$ Voronoi region of A is potentially empty. The *order-k Voronoi diagram* is the collection of all non-empty order-k Voronoi regions. Its *complexity* is the number of such non-empty regions.

[4] For convenience reasons, we sometimes work with $[-0.5, 0.5]^d$ instead of $[0,1]^d$.

[5] We note, in the context of weighted Voronoi diagrams, it is common to only use the normalized weights (just calling them "weights"). In the context of random networks, however, the non-normalized weights are more common. As both notions have their advantages in different situations, we use both.

4 Core Arguments

Here we only briefly discuss the core arguments. See the full version for detailed proofs [12].

4.1 Power-Law SAT. We use a framework that Ben-Sasson and Wigderson [7] introduced for the uniform SAT model. We prove lower bounds for the resolution width, which implies lower bounds for the resolution size and the tree-like resolution size, which then implies lower bounds for the running times of CDCL and DPLL solvers, respectively.

To bound the resolution width, we essentially have to show that different clauses do not overlap too heavily. Specifically, a formula has resolution width $\Omega(w)$ if (1) every set S of at most w clauses contains at least $|S|$ different variables and (2) every set S of $\frac{1}{3}w \leq |S| \leq \frac{2}{3}w$ clauses contains at least a constant fraction of unique variables.

We achieve the bounds in Theorem 2.1 (i–iii) by showing the above two properties directly. For the bound in Theorem 2.1 (iv), we first observe that both properties are fulfilled if the clause-variable incidence graph of a k-CNF formula Φ has high enough bipartite expansion. Recall the definition of bipartite expansion from Section 3 and note how the requirement that the neighborhood of clause vertices is large resembles the requirement that clauses to not overlap too heavily. We show that $G(\Phi)$ is a bipartite expander asymptotically almost surely if Φ is drawn from the power-law model, which yields the lower bound of Theorem 2.1 (iv).

Compared to the uniform case, the weights make the properties required for the lower bounds less likely. Variables with high weight appear in many clauses, making the clauses less diverse. Thus, it is less likely that every clause set covers a large variety of variables.

4.2 Geometric SAT. To explain the core idea of our proof, consider the following simplified geometric model. Map n variables and m clauses to distinct points in the 2-dimensional Euclidean plane (randomly or deterministically). Build the SAT instance by including in each clause c the k variables with the smallest geometric distance to c. Now consider the order-k Voronoi diagram defined by the positions of the n variables. As a clause c contains the k closest variables, the k variables contained in c are exactly the k variables defining the Voronoi region of c's position. Independent of the positions of the n variables, there are only at most $2k\,(n-k)$ regions in the order-k Voronoi diagram [14]. Thus, if we have at least $2^k 2k\,(n-k)$ clauses, then, by the pigeonhole principle, at least one Voronoi region contains 2^k clauses. As k is considered to be a constant, this number of clauses is linear in n, i.e., we still have constant density. Moreover, as repeating the same clause (with the same variable negations) is avoided whenever possible, there is a set of k variables that has a clause for every combination of literals. Thus, we have an unsatisfiable subformula of constant size 2^k, which implies low proof complexity.

This result can be varied and strengthened in multiple ways, e.g., by allowing weighted variables, a higher dimensional ground space, or by softening the requirement that every clause contains the k closest variables (model with higher temperature). In the following, we briefly discuss how these generalizations can be achieved.

Abstract Geometric Spaces. The result by Bohler et al. [14] on the complexity of order-k Voronoi diagrams is very general in the sense that it holds for abstract Voronoi diagrams. Roughly speaking, abstract Voronoi diagrams are based on separating curves between pairs of points that take the role of perpendicular bisectors. In this way, one can abstract from the specific geometric ground space. Whether a point $\boldsymbol{p}$ is closer to site $\boldsymbol{s}_1$ or to site $\boldsymbol{s}_2$ is no longer determined by comparing distances $\|\boldsymbol{s}_1 - \boldsymbol{p}\|$ and $\|\boldsymbol{s}_2 - \boldsymbol{p}\|$ but by the curve separating $\boldsymbol{s}_1$ from $\boldsymbol{s}_2$. For this to work, the separating curves have to satisfy a handful of basic axioms. These are for example satisfied by perpendicular bisectors in the Euclidean or the hyperbolic plane. Thus, the above argumentation for low proof complexity directly carries over to the hyperbolic plane, or more generally, to any abstract geometric space satisfying the axioms.

Lower Density Via Random Clause Positions. Assume the variable positions are fixed. Now choose random positions for the clauses and observe in which regions of the order-k Voronoi diagram they end up. We want to know whether there is a region that contains at least 2^k clauses. This comes down to a balls into bins experiment. Each Voronoi region is a bin and each clause is a ball. Thus, there are $O(n)$ bins and m balls. Moreover, we are interested in the maximum load, i.e., the maximum number of balls that land in a single bin. Due to a result by Raab and Steger [49], the maximum load is a. a. s. in $\Omega(\frac{\log n}{\log\log n})$ if we throw $\Omega(\frac{n}{\text{polylog}\, n})$ balls. Thus, even for a slightly sublinear number of balls, the maximum load is superconstant. We note that this result holds for uniform bins. In our case, we have non-uniform bins, as the probability for a clause to end up in a particular Voronoi region is proportional to the area of the region. However, it is not hard to see that the result by Raab and Steger [49] remains true for non-uniform bins; see the full version [12]. Thus, even if the number of clauses m is slightly sublinear in the number

of variables n, we get a small unsatisfiable subformula asymptotically almost surely if the Voronoi diagram has low complexity.

Positive or Negative Literals with Repetition. Above we assumed that we get the exact same clause with coinciding negations twice only if we already have more than 2^k clauses with the same set of k variables. Although this is arguably a reasonable assumption for the model, we can make a similar argument without it. Assume instead that for each variable, we choose the positive and negative literal uniformly at random, independently of all other choices. Moreover, assume for an increasing function f, that there are $f(n)$ clauses that have the same set of k variables. With the above balls into bins argument, we, e.g., have $f(n) \in \Omega(\frac{\log n}{\log \log n})$. Then the probability that there is a combination of positive and negative literals that we did not see at least once is at most $2^k(1-2^{-k})^{f(n)}$. This probability goes to 0 for $n \to \infty$, i.e., a. a. s., there is an unsatisfiable subformula of constant size 2^k.

Higher Dimension and Weighted Variables. At the core of our argument lies the fact that order-k Voronoi diagrams have linear complexity in the plane. As already mentioned in Section 2.3, this is no longer true for order-k Voronoi diagrams in higher dimensions or if the variables have multiplicative weights. However, for sites distributed uniformly at random, we can show that the complexity can be expected to be linear in the total weight, even in the more general setting. Thus, using that the variables have random positions (a requirement we did not need before), we can apply the above argument to obtain low proof complexity.

Non-Zero Temperature. Non-zero temperatures make it so that clauses do not necessarily contain the k closest variables. Instead, variables are included with probabilities depending on the distance. Thus, we cannot simply look at the order-k Voronoi diagram to determine which variables are contained in a given clause. To resolve this, we call a clause *nice*, if it behaves as it would in the $T = 0$ case, i.e., if it includes the k closest variables. We can show that, in expectation, a constant fraction of clauses is actually nice. Moreover, we can show that the number of nice clauses is concentrated around its expectation. With this, we can apply the same arguments as before to only the nice clauses, of which we have linearly many, to obtain a low proof complexity.

4.3 Voronoi Diagrams. The worst-case lower bounds for the complexity of order-k Voronoi diagrams follow from existing lower bounds on the number of vertices together with Theorem 2.4, which connects the complexity in terms of regions with the complexity in terms of vertices. This connection is obtained by observing how the order-k Voronoi diagram changes when increasing k.

For the average-case linear upper bound on the number of regions, the argument works roughly as follows, assuming the unweighted case for the sake of simplicity. For each size-k subset A of the sites, we devise an upper bound on the probability that A has non-empty order-k Voronoi region. This region is non-empty if and only if there are points that have A as the k closest sites, i.e., if there is a ball that contains the sites of A and no other sites. With this observation, we can use a win-win-style argument. Either the radius of this ball is small, which makes it unlikely that all sites of A lie in the ball, or the ball is large, which makes it unlikely that it contains no other sites.

Acknowledgments. The authors would like to thank Thomas Sauerwald for the fruitful discussions on random SAT models and bipartite expansion.

References

[1] Carlos Ansótegui, Maria Luisa Bonet, and Jordi Levy. On the structure of industrial SAT instances. In *Principles and Practice of Constraint Programming (CP)*, pages 127–141, 2009. doi:10.1007/978-3-642-04244-7_13.

[2] Carlos Ansótegui, Jesús Giráldez-Cru, and Jordi Levy. The community structure of SAT formulas. In *Theory and Applications of Satisfiability Testing (SAT)*, pages 410–423, 2012. doi:10.1007/978-3-642-31612-8_31.

[3] Carlos Ansótegui, María Luisa Bonet, and Jordi Levy. Towards industrial-like random SAT instances. In *International Joint Conference on Artifical Intelligence (IJCAI)*, pages 387–392, 2009. doi:10.5555/1661445.1661507.

[4] Franz Aurenhammer and Herbert Edelsbrunner. An optimal algorithm for constructing the weighted Voronoi diagram in the plane. *Pattern Recognition*, 17(2):251–257, 1984. doi:10.1016/0031-3203(84)90064-5.

[5] Franz Aurenhammer, Rolf Klein, and Der-Tsai Lee. *Voronoi Diagrams and Delaunay Triangulations.* WORLD SCIENTIFIC, 2013. doi:10.1142/8685.

[6] Paul Beame and Ashish Sabharwal. Non-restarting SAT solvers with simple preprocessing can effi-

ciently simulate resolution. In *Conference on Artificial Intelligence (AAAI)*, pages 2608–2615, 2014. URL https://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8397.

[7] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow - resolution made simple. *J. ACM*, 48 (2):149–169, 2001. doi:10.1145/375827.375835.

[8] Marcin Bienkowski, Valentina Damerow, Friedhelm Meyer auf der Heide, and Christian Sohler. Average case complexity of Voronoi diagrams of n sites from the unit cube. In *European Workshop on Computational Geometry (EuroCG)*, pages 167–170, 2005. URL http://www.win.tue.nl/EWCG2005/Proceedings/43.pdf.

[9] Thomas Bläsius, Cedric Freiberger, Tobias Friedrich, Maximilian Katzmann, Felix Montenegro-Retana, and Marianne Thieffry. Efficient shortest paths in scale-free networks with underlying hyperbolic geometry. In *International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 20:1–20:14, 2018. doi:10.4230/LIPIcs.ICALP.2018.20.

[10] Thomas Bläsius, Philipp Fischbeck, Tobias Friedrich, and Maximilian Katzmann. Solving vertex cover in polynomial time on hyperbolic random graphs. In *Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 154, pages 25:1–25:14, 2020. doi:10.4230/LIPIcs.STACS.2020.25.

[11] Thomas Bläsius, Tobias Friedrich, and Andrew M. Sutton. On the empirical time complexity of scale-free 3-SAT at the phase transition. In *Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, pages 117–134, 2019. doi:10.1007/978-3-030-17462-0_7.

[12] Thomas Bläsius, Tobias Friedrich, Andreas Göbel, Jordi Levy, and Ralf Rothenberger. The impact of heterogeneity and geometry on the proof complexity of random satisfiability. *CoRR*, abs/2004.07319, 2020. URL https://arxiv.org/abs/2004.07319.

[13] Michel Bode, Nikolaos Fountoulakis, and Tobias Müller. On the giant component of random hyperbolic graphs. In *European Conference on Combinatorics, Graph Theory and Applications*, pages 425–429, 2013. doi:10.1007/978-88-7642-475-5_68.

[14] Cecilia Bohler, Panagiotis Cheilaris, Rolf Klein, Chih-Hung Liu, Evanthia Papadopoulou, and Maksym Zavershynskyi. On the complexity of higher order abstract voronoi diagrams. *Computational Geometry: Theory and Applications*, 48(8): 539–551, 2015. doi:10.1016/j.comgeo.2015.04.008.

[15] Jean-Daniel Boissonnat, Micha Sharir, Boaz Tagansky, and Mariette Yvinec. Voronoi diagrams in higher dimensions under certain polyhedral distance functions. *Discrete & Computational Geometry*, 19 (4):485–519, 1998. doi:10.1007/PL00009366.

[16] Barry N. Boots. Weighting thiessen polygons. *Economic Geography*, 56(3):248–259, 1980. doi:10.2307/142716.

[17] Karl Bringmann, Ralph Keusch, and Johannes Lengler. Sampling geometric inhomogeneous random graphs in linear time. In *Annual European Symposium on Algorithms (ESA)*, volume 87, pages 20:1–20:15, 2017. doi:10.4230/LIPIcs.ESA.2017.20.

[18] Fan Chung and Linyuan Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, 2002. doi:10.1073/pnas.252631999.

[19] Fan Chung and Linyuan Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6(2):125–145, 2002. doi:10.1007/PL00012580.

[20] Vasek Chvátal and Endre Szemerédi. Many hard examples for resolution. *J. ACM*, 35(4):759–768, 1988. doi:10.1145/48014.48016.

[21] Bogdan Cundefinedrbunar, Ananth Grama, Jan Vitek, and Octavian Cundefinedrbunar. Redundancy and coverage detection in sensor networks. *ACM Transactions on Sensor Networks*, 2(1):94–128, 2006. doi:10.1145/1138127.1138131.

[22] Elena Deza and Michel-Marie Deza. *Chapter 20 - Voronoi Diagram Distances*, pages 253–261. Elsevier, 2006. doi:10.1016/B978-044452087-6/50020-2.

[23] Anne Driemel, Sariel Har-Peled, and Benjamin Raichel. On the expected complexity of Voronoi diagrams on terrains. *ACM Transactions on Algorithms*, 12(3), 2016. doi:10.1145/2846099.

[24] Rex A. Dwyer. Higher-dimensional Voronoi diagrams in linear expected time. *Discrete & Computational Geometry*, 6(3):343–367, 1991. doi:10.1007/BF02574694.

[25] Jeff Erickson. Nice point sets can have nasty Delaunay triangulations. In *Annual Symposium on Computational Geometry (SoCG)*, pages 96–105, 2001. doi:10.1145/378583.378636.

[26] Jeff Erickson. Dense point sets have sparse Delaunay triangulations: Or "... but not too nasty". *Discrete & Computational Geometry*, 33:83–115, 2005. doi:10.1007/s00454-004-1089-3.

[27] Chenglin Fan and Benjamin Raichel. Linear expected complexity for directional and multiplicative voronoi diagrams. In *Annual European Symposium on Algorithms (ESA)*, pages 45:1–45:18, 2020. doi:10.4230/LIPIcs.ESA.2020.45.

[28] Tobias Friedrich and Ralf Rothenberger. The satisfiability threshold for non-uniform random 2-SAT. In *International Colloquium on Automata, Languages, and Programming, (ICALP)*, volume 132 of *LIPIcs*, pages 61:1–61:14, 2019. doi:10.4230/LIPIcs.ICALP.2019.61.

[29] Tobias Friedrich, Anton Krohmer, Ralf Rothenberger, Thomas Sauerwald, and Andrew M. Sutton. Bounds on the satisfiability threshold for power law distributed random SAT. In *Annual European Symposium on Algorithms (ESA)*, pages 37:1–37:15, 2017. doi:10.4230/LIPIcs.ESA.2017.37.

[30] Tobias Friedrich, Anton Krohmer, Ralf Rothenberger, and Andrew M. Sutton. Phase transitions for scale-free SAT formulas. In *Conference on Artificial Intelligence,(AAAI)*, pages 3893–3899, 2017. URL `http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14755`.

[31] Lauro C. Galvão, Antonio G.N. Novaes, J.E. Souza de Cursi, and João C. Souza. A multiplicatively-weighted Voronoi diagram approach to logistics districting. *Computers & Operations Research*, 33 (1):93–114, 2006. doi:10.1016/j.cor.2004.07.001.

[32] Andreas Gemsa, D. T. Lee, Chih-Hung Liu, and Dorothea Wagner. Higher order city Voronoi diagrams. In *Scandinavian Workshop on Algorithm Theory (SWAT)*, pages 59–70, 2012. doi:10.1007/978-3-642-31155-0_6.

[33] Jesús Giráldez-Cru and Jordi Levy. A modularity-based random SAT instances generator. In *International Joint Conference on Artifical Intelligence (IJCAI)*, pages 1952—-1958, 2015. doi:10.5555/2832415.2832520.

[34] Jesús Giráldez-Cru and Jordi Levy. Description of popularity-similarity SAT instances. In Tomáš Balyo, Marijn J. H. Heule, and Matti Järvisalo, editors, *Proceedings of SAT Competition 2017: Solver and Benchmark Descriptions*, pages 49–50, 2017. URL `https://helda.helsinki.fi/handle/10138/224324`.

[35] Jesús Giráldez-Cru and Jordi Levy. Locality in random SAT instances. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 638–644, 2017. doi:10.24963/ijcai.2017/89.

[36] Mordecai J. Golin and Hyeon-Suk Na. On the average complexity of 3D-Voronoi diagrams of random points on convex polytopes. *Computational Geometry*, 25(3):197–231, 2003. doi:10.1016/S0925-7721(02)00123-2.

[37] Sariel Har-Peled and Benjamin Raichel. On the complexity of randomly weighted multiplicative Voronoi diagrams. *Discrete & Computational Geometry*, 53(3):547–568, 2015. doi:10.1007/s00454-015-9675-0.

[38] Victor Klee. On the complexity of d-dimensional Voronoi diagrams. *Archiv der Mathematik*, 34(1): 75–80, 1980. doi:10.1007/BF01224932.

[39] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguñá. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, 2010. doi:10.1103/PhysRevE.82.036106.

[40] Ngọc-Minh Lê. On Voronoi diagrams in the L_p-metric in $\mathbb{R}^d$. *Discrete & Computational Geometry*, 16(2):177–196, 1996. doi:10.1007/BF02716806.

[41] Der-Tsai Lee. On k-nearest neighbor Voronoi diagrams in the plane. *IEEE Transactions on Computers*, 31:478–487, 1982. doi:10.1109/TC.1982.1676031.

[42] Chih-Hung Liu, Evanthia Papadopoulou, and D. T. Lee. An output-sensitive approach for the L_1/L_∞ k-nearest-neighbor Voronoi diagram. In *Annual European Symposium on Algorithms (ESA)*, pages 70–81, 2011. doi:10.1007/978-3-642-23719-5_7.

[43] Nathan Mull, Daniel J. Fremont, and Sanjit A. Seshia. On the hardness of SAT with community structure. In *Theory and Applications of Satisfiability Testing (SAT)*, pages 141–159, 2016. doi:10.1007/978-3-319-40970-2_10.

[44] Ketan Mulmuley. On levels in arrangements and Voronoi diagrams. *Discrete & Computational Geometry*, 6(3):307–338, 1991. doi:10.1007/BF02574692.

[45] Tobias Müller and Merlijn Staps. The diameter of KPKVB random graphs. *Advances in Applied Probability*, 51(2):358–377, 2019. doi:10.1017/apr.2019.23.

[46] Zack Newsham, Vijay Ganesh, Sebastian Fischmeister, Gilles Audemard, and Laurent Simon. Impact of community structure on SAT solver performance. In Carsten Sinz and Uwe Egly, editors, *Theory and Applications of Satisfiability Testing (SAT)*, pages 252–268, 2014. doi:10.1007/978-3-319-09284-3_20.

[47] Fragkiskos Papadopoulos, Maksim Kitsak, M. Ángeles Serrano, Marián Boguñá, and Dmitri Krioukov. Popularity versus similarity in growing networks. *Nature*, 489:537–540, 2012. doi:10.1038/nature11459.

[48] Knot Pipatsrisawat and Adnan Darwiche. On the power of clause-learning SAT solvers as resolution engines. *Artificial Intelligence*, 175(2):512–525, 2011. doi:10.1016/j.artint.2010.10.002.

[49] Martin Raab and Angelika Steger. "Balls into bins" — a simple and tight analysis. In *Randomization and Approximation Techniques in Computer Science*, pages 159–170, 1998. doi:10.1007/3-540-49543-6_13.

[50] Raimund Seidel. On the number of faces in higher-dimensional Voronoi diagrams. In *Annual Symposium on Computational Geometry (SoCG)*, pages 181–185, 1987. doi:10.1145/41958.41977.

[51] Michael Ian Shamos and Dan Hoey. Closest-point problems. In *Annual Symposium on Foundations of Computer Science (FOCS)*, pages 151–162, 1975. doi:10.1109/SFCS.1975.8.

[52] Moshe Y. Vardi. Boolean satisfiability: Theory and engineering. *Communications of the ACM*, 57(3): 5–5, 2014. doi:10.1145/2578043.

[53] Ivan Voitalov, Pim van der Hoorn, Remco van der Hofstad, and Dmitri Krioukov. Scale-free networks well done. *Physical Review Research*, 1:033034, 2019. doi:10.1103/PhysRevResearch.1.033034.

[54] Edward Zulkoski, Ruben Martins, Christoph M. Wintersteiger, Robert Robere, Jia Liang, Krzysztof Czarnecki, and Vijay Ganesh. Relating complexity-theoretic parameters with SAT solver performance. *CoRR*, abs/1706.08611, 2017. URL `http://arxiv.org/abs/1706.08611`.

[55] Edward Zulkoski, Ruben Martins, Christoph M. Wintersteiger, Jia Hui Liang, Krzysztof Czarnecki, and Vijay Ganesh. The effect of structural measures and merges on SAT solver performance. In *Principles and Practice of Constraint Programming (CP)*, pages 436–452, 2018. doi:10.1007/978-3-319-98334-9_29.