Testing for Geometry in Random Satisfiability Instances

Motivation SAT is the archetypal NP-complete problem and is used as a starting point for hardness reductions in theory, but in practice heuristic SAT solving algorithms can solve large-scale industrial SAT instances very efficiently. One possible explanation for this disparity is that industrial SAT instances have particular properties that make them more easily solvable.

One approach to understanding this difference between theory and practice is to study random SAT models whose generated instances share structural properties with real-world SAT instances. This approach has been successfully applied in network problems, where *geometric random graph* models have been shown to share many properties with real-world networks, be it structural or algorithmic.

Two industrial SAT instances (left and middle) and a geometric SAT instance (right).

In a recent paper [2] we studied how two properties of random SAT instances affect their tractability. More importantly we showed that when SAT instances are generated with a particular underlying geometry, they can be easily refuted, while this is not the case when the geometry is removed. While in the non-geometric mode each variable is added to each clause independently according to its weight, the geometric model affects the correlation between its clauses as follows. Variables and clauses are dropped uniformly at random as points in a compact space, and then each clause connects to its *k* closest variables. This way clauses that are geometrically close are likely to contain the same variables.

While this is significant progress in explaining the tractability of SAT instances, it does not provide us with a complete answer. The reason is that this geometric random SAT model generates instances that are "too easy" compared to the real-world ones. That is, generated SAT formulae have constant-size unsatisfiable subformulae, while real-world instances have logaritmic-size ones.

Goal of the project The goal of the project is to study the structural properties of geometric SAT instances and compare them to those of real world instances. This could be done in several ways.

1. Obtaining bounds on the total variation distance between the geometric random k-SAT distribution and some basic/non-geometric generation procedure, such as uniform random connection.
2. Developing (with proof) a stastical test for whether a given instance has underlying geometry. Possible test statistics could be degree distribution, clustering coefficient, diameter, etc.
3. Experimentally finding a test statistic for differentiating between real-world SAT instances and geometric random k-SAT instances; for recent examples see [1, 3].

It would also be interesting to understand how any of these analyses depend on the dimension of the geometric space.

What we expect from you We expect students who are familiar with graph theory and algorithms, enjoy writing proofs, and interested in learning more probabililty theory. We understand that you won't know all the background and we'll provide you with sources for learning. You should bring the willingness and curiosity to delve into an interesting interdisciplinary research topic in theoretical computer science, which makes use of tools from several domains of mathematics.

What you can expect from us We will gently introduce you to the field and accompany you all along this interesting journey. This will be a team effort, and we aim at publishing our results at a renowned international conference.

How to contact us You're welcome to visit us in K2 or send us an e-mail:

Prof. Dr. Tobias Friedrich ⎫
Dr. Andreas Göbel ⎬ firstname.lastname@hpi.de
Dr. Samuel Baguley ⎭

[1] P. Almagro, M. Boguña, M. Ángeles Serrano (2022): *Detecting the ultra low dimensionality of real networks*. DOI: https://doi.org/10.1038/s41467-022-33685-z.
[2] T. Bläsius, T. Friedrich, A. Göbel, J. Levy, R. Rothenberger (2020): *The impact of heterogeneity and geometry on the proof complexity of random satisfiability.* DOI: https://doi.org/10.1137/1.9781611976465.4. Arxiv: https://arxiv.org/abs/2004.07319.
[3] T. Friedrich, A. Göbel, M. Katzmann, L. Schiller (2023): *A simple statistic for determining the dimensionality of complex networks*. Arxiv: https://arxiv.org/abs/2302.06357.