Phase transition and finite-size scaling in the vertex-cover problem

Alexander K. Hartmann [a,*] Wolfgang Barthel [a] Martin Weigt [b]

[a] *Institut für Theoretische Physik, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany*
[b] *ISI Foundation, Viale S. Severo 65, 10133 Torino, Italy*

Abstract

NP-complete problems play a fundamental role in theoretical computer science. Recently, phase transitions were discovered in such problems, when studying suitably parameterized random ensembles. By applying concepts and methods from statistical physics, it is possible to understand these models and the phase transitions which occur. Here, we consider the vertex-cover problem, one of the six "basic" NP-complete problems. We describe the phase transition and the running time of an exact algorithm around the phase transition. To investigate how this transition resembles a phase transition in physical systems, we apply finite-size scaling and we study a correlation-length like quantity.

Key words: Combinatorial Optimization, Phase Transitions, NP-complete
PACS: 64.60.-i, 05.20.-y, 02.10.Eb

1 Introduction

During the last three decades, more and more problems in physics have been solved by applying optimization algorithms [1]. This development is mainly due to the strong increase in computer power and because of the rapid developments of new sophisticated optimization algorithms. This means that physicists are applying tools from computer science to study physics problems. Recently, knowledge has been transferred also from physics to computer science. One famous example is the invention of the simulated annealing algorithm [2]. It allows to solve optimization problems heuristically, by interpreting the function to be optimized as an energy, introducing an artificial temperature and cooling the system gradually using a Monte Carlo simulation.

* Corresponding Author:
Email address: hartmann@physik.uni-goe.de (Alexander K. Hartmann).

More recently, concepts and methods from statistical physics have been applied to study "classical" problems from theoretical computer science. The most interesting problems are those from the class of *NP-complete* problems [3]. For these problems only algorithms are known, which have a *worst case* running time which increases exponentially with the size of the problems, hence one calls these problems *hard*. This means no fast algorithm is available, i.e. an algorithm having a running time which grows only polynomially (P) with the system size. On the other hand, up to now nobody has been able to prove that no polynomial algorithm exists, i.e. it might exist. This is the famous *P-NP problem*. In order to solve the P-NP problem, one has to understand the source of the apparent hardness of such problems.

However, in many cases one is interested in a (somehow defined) *typical* behavior, instead of the worst case behavior. To elucidate typical cases, one has started to study suitably parametrized random ensembles of problems. For these ensembles, phase transitions in the solvability have been observed [4], and the hardest to solve instances are often found right at the phase boundaries. Hence, to understand the source of the hardness, one can study the problems occurring close to phase transitions. In this context, also analytical tools borrowed from statistical mechanics are very valuable. Combinatorial problems are mapped to equivalent physical models, and methods like the replica trick or the cavity approach [5] are used to calculate and analyze phase diagrams. One main observation is that computational hardness is typically related to the glassiness of the corresponding physical system. Furthermore, the knowledge of the phase diagram, allows to calculate the typical running time of an exact algorithm analytically [6,7]. Very recently, very efficient algorithms have been invented on the basis of statistical mechanics approaches, namely the *survey propagation* algorithm [8].

Here, to introduce to this field and to illustrate some crucial phenomena, we present the vertex-cover problem. In the next section, we define the problem. Then, we demonstrate the occurrence of a phase transition and the corresponding change in the running time of a vertex-cover algorithm. Finally, we present new results on the usefulness of critical exponents to describe the phase transition and on the existence of a correlation-length like quantity diverging at this phase transition.

2 The vertex-cover problem

Let $G = (V, E)$ be a graph, V being the set of vertices and $E \subset V^2$ the set of undirected edges. In particular we study an ensemble of Erdös-Rényi random graphs [9], where each graph has $N = |V|$ vertices and $M = cN/2$ randomly selected edges, c being the average connectivity of single vertices. Note that at $c = 1$ a percolation transition occurs.

Let be $V' \subset V$ a subset of vertices of size $X = |V'|$. We call an *edge* $\{i, j\} \in E$

covered, iff $i \in V'$ or $j \in V'$. We call the *graph* G covered iff *all* edges are covered by V' and in this case V' is a vertex cover. The *vertex-cover problem* (VC) is a decision problem, i.e. given by a question whose answer can be only "*yes*" or "*no*". For VC, the question is: *For given graph G and given value X, exists there a vertex cover of size X?* A related optimization problem is: for given graph G and given value X, what is the minimum *energy*, which is defined as the number of uncovered edges? Usually, for the optimization type of problem there are many solutions, i.e. the problem is degenerate. Typically, the number of solutions grows exponentially with system size. Note that VC has been shown to be *NP-complete* [3], hence it plays a central role in computational complexity.

3 The phase transition

One can use an exact algorithm [10] to solve the related optimization problem. If the minimum number of uncovered edges is zero, the graph is coverable by X vertices, i.e. the answer to the decision problem is "yes".

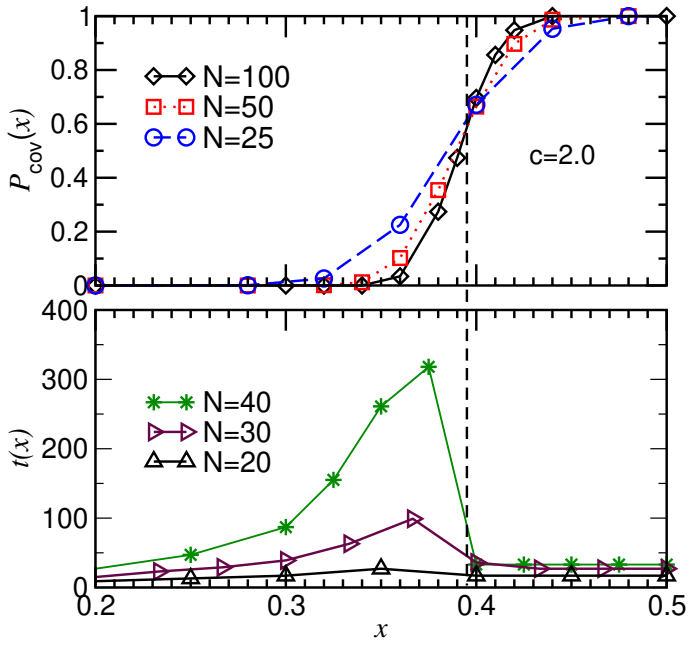

Fig. 1. Top: Probability that a random graph of average connectivity $c = 2$ has a vertex cover of relative size $x = X/N$. Bottom: Running time of an exact algorithm to solve VC with given value of x.

In Fig. 1, we show the numerically determined probability that a random graph with $c = 2.0$ is coverable using a fraction $x = X/N$ of all vertices. One observes a strong increase near $x_c = 0.3919$. This increase sharpens for larger N, and all curves cross close to x_c. Hence, there appears to be a *phase transition* in the solvability of VC at a critical value x_c. Furthermore, the running time of the optimization algorithm

is shown. One can clearly see a peak close to the phase transition, i.e. there the hardest instances are located.

One can solve the problem for different values c of the connectivity. The function $x_c(c)$ determines the phase boundary of the problem. For small connectivities $c \leq e \approx 2.71828$, the phase boundary can be calculated [10] exactly in the replica-symmetric framework. We have studied [11] the structure of the exponentially large solution space, right at the phase boundary, i.e. minimum vertex covers. For this, one considers minimum solutions V', V'' as being *neighbors*, *iff* they have the same size and differ only by one element. *Clusters* are defined as the transitive closures of the neighbor relation. For these small connectivities $c \leq e$, all solutions are usually collected in one cluster, representing the replica-symmetric nature of the analytical solution. For larger connectivities replica symmetry is only an approximation and many solution clusters exist, as numerically found [11]. Interestingly, one can use the phase diagram to calculate analytically the running time of a simple version of the optimization algorithm [7].

4 Finite-size scaling

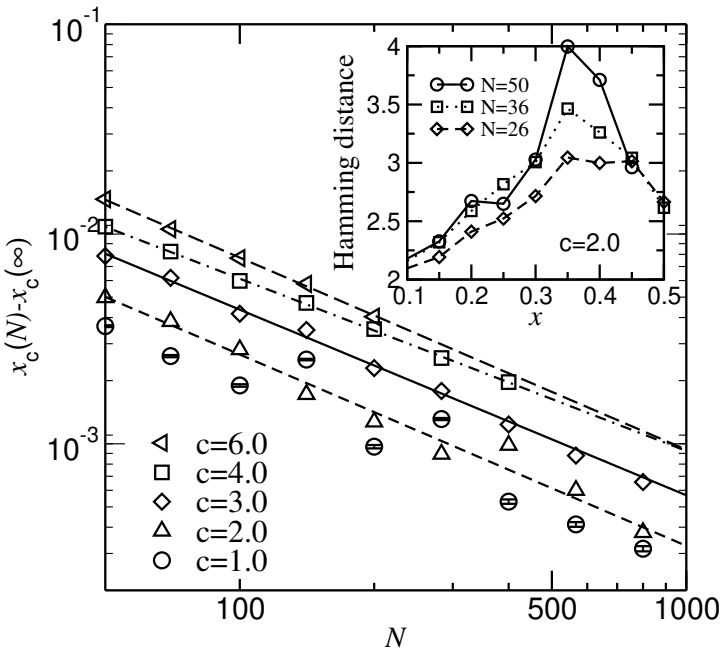

Fig. 2. Finite-size scaling behavior of the critical cover size. The inset shows the scaling of the correlation volume as a function of x for different sizes. Error bars are at most of the order of the symbol size.

To elucidate whether the cov-uncov transition has more in common with a physical phase transition, we study the finite-size scaling behavior of the critical value $x_c(N)$ as a function of system size, by fitting it to functions of the form $x_c(\infty) + aN^{-b}$.

Such behavior corresponds to the scaling for the critical temperature T_c near second order phase transitions [12], where the critical exponent ν of the correlation length determines the shift of T_c: $T_c(L) = T_c(\infty) + aT_c(L)^{-1/\nu}$. And indeed, for $c > 1$, we find a good scaling behavior, see Fig. 2, while for $c = 1$, at the percolation transition, the data scatters strongly. For $c > 1$, we get exponents $b(c = 2) = 0.91(9)$, $b(3) = 0.88(4)$, $b(4) = 0.82(4)$, and $b(6) = 0.92(11)$. Hence, within error bars, this exponent seems to be universal.

One is tempted to ask: is this scaling behavior connected to a divergence of a correlation length-like quantity, as in conventional second order phase transitions? Since VC is a mean-field model, it does not exhibit spatial length scales like finite-dimensional systems, so we cannot introduce a correlation length. Instead, we study the *correlation volume*, which we define as the minimum number of vertices which change, if we force *one* randomly selected vertex to change its state *covered* $\leftrightarrow$ *uncovered* in a ground state. Numerically this means we do the following thing:

- We enumerate all ground states.
- We choose one ground state V' randomly.
- We choose one vertex i_0 randomly among the not *frozen* vertices. I.e. the vertex is neither in *all* ground states *covered*, nor in *all* ground states *uncovered*.
- Among all other ground states, we pick that one V'', which is closest to V' given that is state of i_0 is different in V' and V''
- The number of vertices which are different in V' and V'', i.e. $|(V' \setminus V'') \cup (V'' \setminus V')|$ is the correlation volume.

Indeed, this correlation volume exhibits a peak near the phase transition, see Fig. 2. Unfortunately, calculating the correlation volume requires to enumerate all ground states, so we cannot treat large systems. Hence, using small sizes, we are not able to draw a precise connection to the exponent $b \approx 0.86$ measured above.

5 Aknowledgements

We thank E.O. Yewande for critically reading the manuscript. Financial support was obtained from the *VolkswagenStiftung* (Germany) within the program "Nachwuchsgruppen an Universitäten".

References

[1] A. K. Hartmann, H. Rieger, Optimization Algorithms in Physics, Whiley-VCH, Berlin, 2001.

[2] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671.

[3] M. R. Garey, D. S. Johnson, Computers and intractability, Freeman, San Francisco, 1979.

[4] S. Kirkpatrick, B. Selman, Critical behavior in the satisfiability of random boolean expressions, Science 264 (1994) 1297.

[5] M. Mézard, G. Parisi, M. A. Virasoro, Spin glass theory and beyond, World Scientific, Singapore, 1987.

[6] S. Cocco, R. Monasson, Trajectories in phase diagrams, growth processes, and computational complexity: How search algorithms solve the 3-satisfiability problem, Phys. Rev. Lett. 86 (2001) 1654.

[7] M. Weigt, A. K. Hartmann, The typical-case complexity of a vertex-covering algorithm on finite-connectivity random graphs, Phys. Rev. Lett. 86 (2001) 1658.

[8] M. Mézard, G. Parisi, R. Zecchina, Analytic and algorithmic solution of random satsifiability problems, Science 297 (2002) 812.

[9] B. Bollobas, Random Graphs, Academic Press, 1985.

[10] M. Weigt, A. K. Hartmann, The number guards needed by a museum – a phase transition in vertex covering of random graphs, Phys. Rev. Lett. 84 (2000) 6118.

[11] W. Barthel, A. K. Hartmann, Clustering analysis of the ground-state structure of the vertex-cover problem, Phys. Rev. E 70 (2004) 066120.

[12] M. N. Barber, in: C. Domb, J. L. Lebowitz (Eds.), Phase Transitions and Critical Phenomena, Vol. 8, Academic Press, London, 1983, p. 146.