Title: Where the Really Hard Problems Are

Author: Cheeseman, Kanefsky & Taylor

Proc.: IJCAI

Pages: 331–337

Year: 1991

Theory and Practice of Constraint Satisfaction

CSCE 990-05, Fall 1999

www.cse.unl.edu/~choueiry/CSE990-05/

Reality

Most **NP**-complete problems are easy to solve $\begin{cases} \text{randomly generated} \\ \text{real-world problems} \end{cases}$

Theory

NP complexity is worst-case analysis for a class of problems
says nothing about difficulty of typical instances

Question

Where are the worst cases?

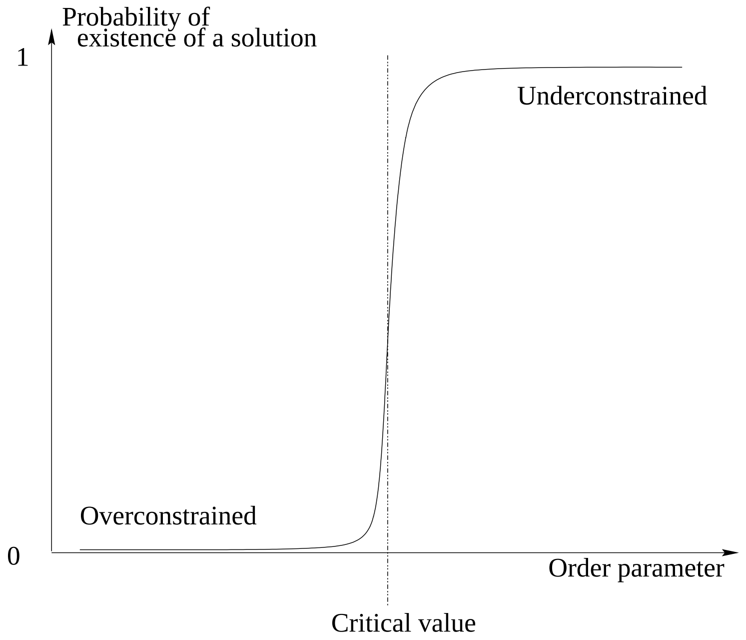
Can we find subclasses of problems that are typically hard to solve?

Answer

Identify 1 or more order parameter for some problems to predict problem difficulty

Order parameter & phase transition (I)

- Hard instances occur around critical value of order parameter
- Critical value separates solvable and unsolvable (easy) instances
- At critical value, mix of 'water and ice'



$\longrightarrow$ Let's produce phase diagrams for practical applications

Order parameter and phase transition (II)

- Describe the behavior of a complex system at the macroscopic level:
  Large number of interactions produces coordinated macroscopic behavior, independent of the detailed interactions within the system

- Characterize a system as a whole

- Summarize the properties of a system through a small set of parameters

Contributions

- Study of the phase transition in $\begin{cases} \text{Hamiltonian circuits} \\ \text{Graph coloring} \\ K\text{-satisfiability} \\ \text{TSP (minimization)} \end{cases}$

- Mapping between problems preserves the critical boundary

- Two conjectures:
  - All **NP**-complete problems have at least one order-parameter
  - **P** problems do not have a phase transition, or, if they do, it occurs for bounded N and thus has a bounded cost

Hamiltonian circuit

Definition: HC is a cyclic ordering of the nodes in a graph

- cyclic: circuit is closed
- all nodes are included
- no node is repeated (circuit does not cross)

Problem: Given a graph, is there a Hamiltonian circuit?

Question 1: How does the <u>probability</u> of existence of a HC varies with the average connectivity (degree of nodes) of the graph?

Answer: by empirical studies on random graphs
- Average connectivity: 0 — 16
- Number of nodes: 8, 16 & 24

Include Figure 1 (a) from paper

- Complete graph: any ordering is a HC
- High average connectivity: high probability of HC, many HCs, easy to find
- Low average connectivity: barely connected, unlikely a HC exist, easy to find
- Theoretical phase transition: $\ln N + \ln \ln N$

Question 2: How does the computational cost of finding a HC varies with the average connectivity (degree) of the graph?

Answer: using backtrack search with 2 heuristics

Include Figure 1 (b)/(c) from paper

- Max threshold is chosen. Log scale.
- Sharpness of phase transition increases with N
- Phase transition in cost occurs at the same point where probability of solution drops

Observations:

- There exists a phase transition around a critical value of the average connectivity (order parameter)
- Order critical connectivity increases with N
- Order critical connectivity separates soluble (high connectivity) and unsoluble (low connectivity) instances
- Intuitive interpretation (for backtrack search)

Note:

Most HC problems are hard and most k-coloring are easy:

HC: ‘global’ constraint; K-colorability: local constraints

So, is the phase transition in HC a result of global constraint?

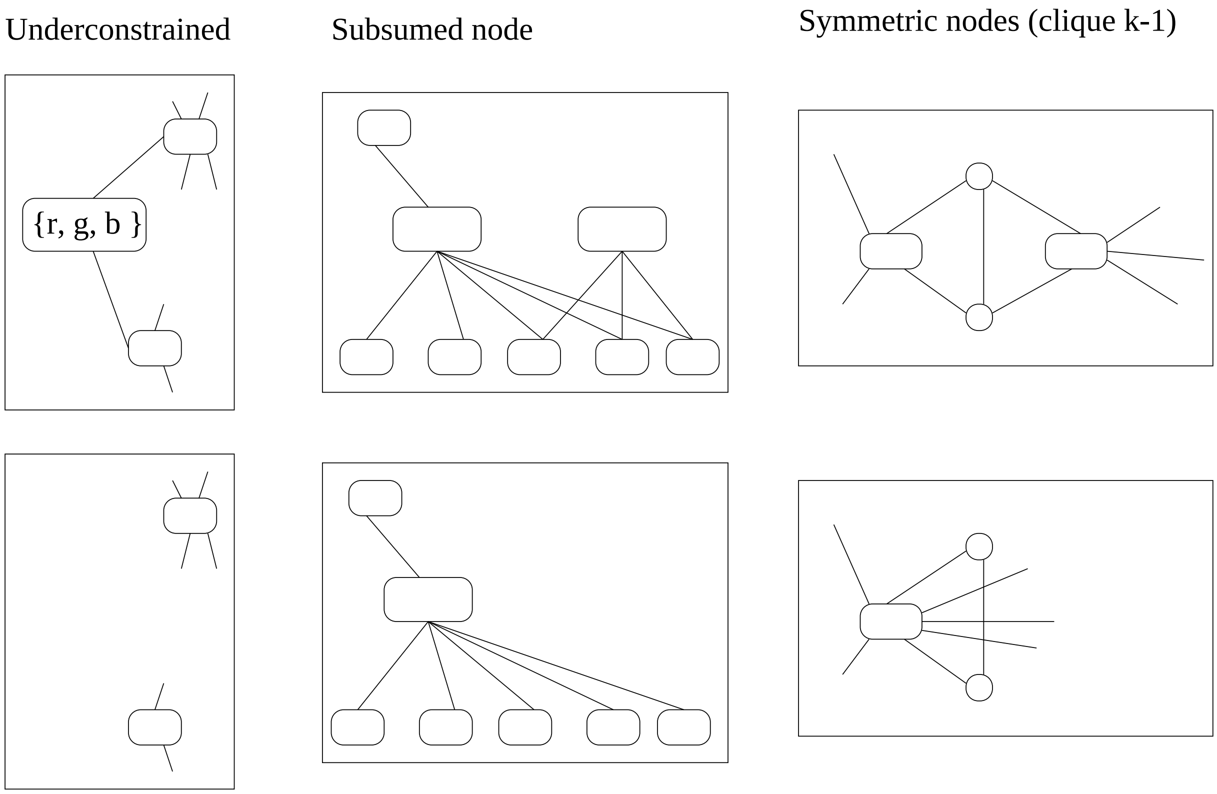
K-colorability

- Decision problem: coloring a graph with K colors
- (Optimization problem: minimize K, chromatic number)
- Note: color rotation symmetry
- Very useful to model practical applications
- Random graphs: generally easy
- First reduce the graph

  3 graph-reduction operators: $\begin{cases} \text{Underconstrained node} \\ \text{Subsumed node} \\ \text{Symmetric nodes} \end{cases}$

- Order parameter: average connectivity (= average constraints per variable)

Reduction operators



- Operators are applied iteratively.. cascade effect.. only tough problems remain
- Reduction preserves solvability and unsolvability
- But total number of solutions is reduced

Existence of solution (Q1)

Include Figure 3 (a) from paper

- The larger K, the higher the critical value
- The larger N, the shaper the transition

Cost of finding a solution (Q2)

Problems generated randomly, but guaranteed soluble

Solved with backtrack search (Brélaz)

Include Figure 3 (b)/(c) from paper

- The phase transition is clear in both cases
- Same location as in the solution-probability transition
- The transition in 4-color is sharper than 3-color

$\longrightarrow$ Previous authors found most coloring problems easy to solve, because they used non-reduced problem instances

K-satisfiability

- Reduced using resolution
- Solved using backtrack search with FFP heuristic
- Order parameter: average constraints per variable
- Same phenomena as before were observed.
- 2-SAT (**P**) has no phase boundary
- Large K-sat ($k \geq 4$) seem to be easy and the transition becomes weaker
  Possible interpretation: reduction is not effective enough, hard problems are diluted by trivial problems

TSP (HC is a special case)

- Cost: matrix representation, in general not symmetric
- Experiments: cost matrix with mean edge cost of 10
  vary standard deviation of costs
- Minimization problem: Q1 cannot be investigated
- Q2: solved by Little's backtrack search, sound and complete
- Order parameter: normalized path standard deviation
- Vertical axis: log scale
- Magnitude and sharpness of phase transition increases with # of cities
- Authors provide intuitive interpretation of results

Mappings between problems

- **Question**:
  Mapping between problems preserves solutions
  How does it affect the phase transition?

  Answer:
  Yes (experimentally), critical boundary maps consistently

  .. warning: mapping does not preserve 'reduction'
  (i.e., the mapping of a reduced problem is not necessarily reduced)

- **Question**:
  How about **P** problems? Do they have a phase transition.

  Answer
  Yes, but the boundary is fixed (does not increase with N)

Research issues (need to double-check)

- It is believed that the order parameter and the phase transition are an intrinsic property of a problem, independent of the backtrack algorithm used to solve it
(some backtrack strategies may significant effects though)
- Some problems may exhibit more than one phase transition

Special issue of the AIJ (Vol 81, 1996)

Warning

Studies use median values, and probabilistic values:

- A problem in the phase transition can be very easy
- A problem in the easy region may be very very difficult
- Problems are randomly generated, and ‘uniform’, are not representative of real applications