Calculation of typical running time of a branch-and-bound algorithm for the vertex-cover problem

Joni Pajarinen, Joni.Pajarinen@iki.fi

October 21, 2007

Alexander K. Hartmann and Martin Weigt, *Phase Transitions in Combinatorial Optimization Problems*, WILEY-VCH (2005)

Martin Weigt and Alexander K. Hartmann, Typical Solution Time for a Vertex-Covering Algorithm on Finite-Connectivity Random Graphs, *Physical Review Letters*, **86**, 8 (2001)

Content

- Vertex-cover (VC) problem
- A branch-and-bound algorithm
- Experimental results
- First-moment method
- Analysis of first descent into the configuration tree
- Backtracking analysis
- Summary

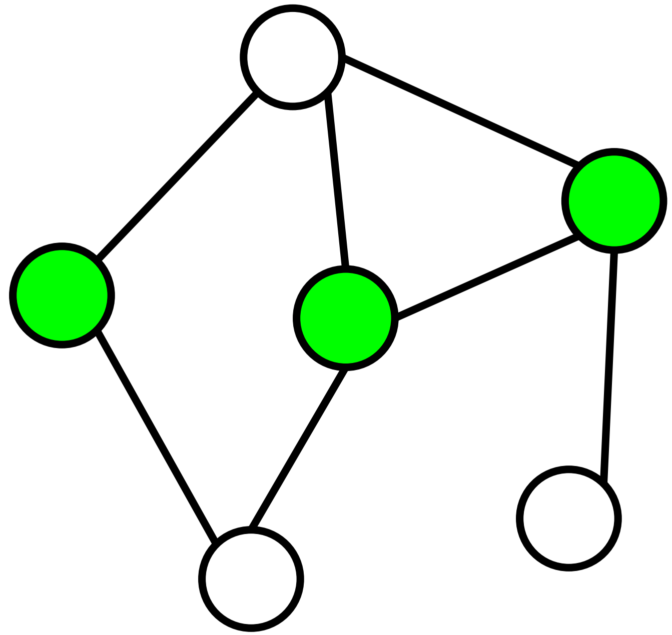
Vertex-cover problem

- Cover vertices so that all edges connect to at least one covered vertex

- Graph: $G = (V, E)$
  - Vertex-cover: $V_{vc} \subseteq V : i \in V_{vc} \vee j \in V_{vc}, \forall (i,j) \in E$

- Minimum vertex-cover: $\underset{V_{vc}}{\arg\min} |V_{vc}|$

- $N = |V|$, xN = size of the largest allowed vertex-cover



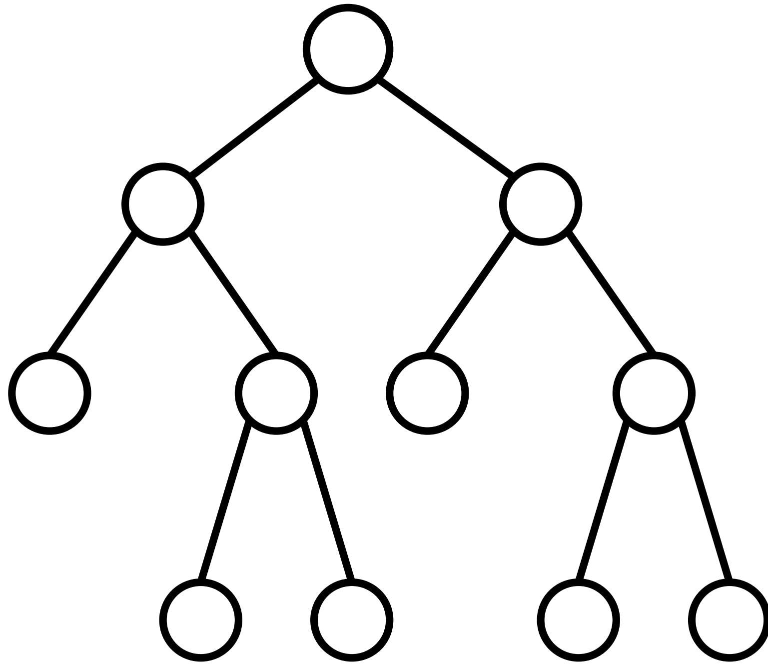
A branch-and-bound vertex-cover algorithm

- Vertice states: free, covered or uncovered. Start with free vertices

- Branch:

  1. Cover a random free vertex, if free or uncovered neighbours exist
  2. Backtrack if we exceed the vertex limit xN
  3. Uncover the free vertex otherwise

- Bound: Don't mark a vertex with an uncovered neighbour uncovered



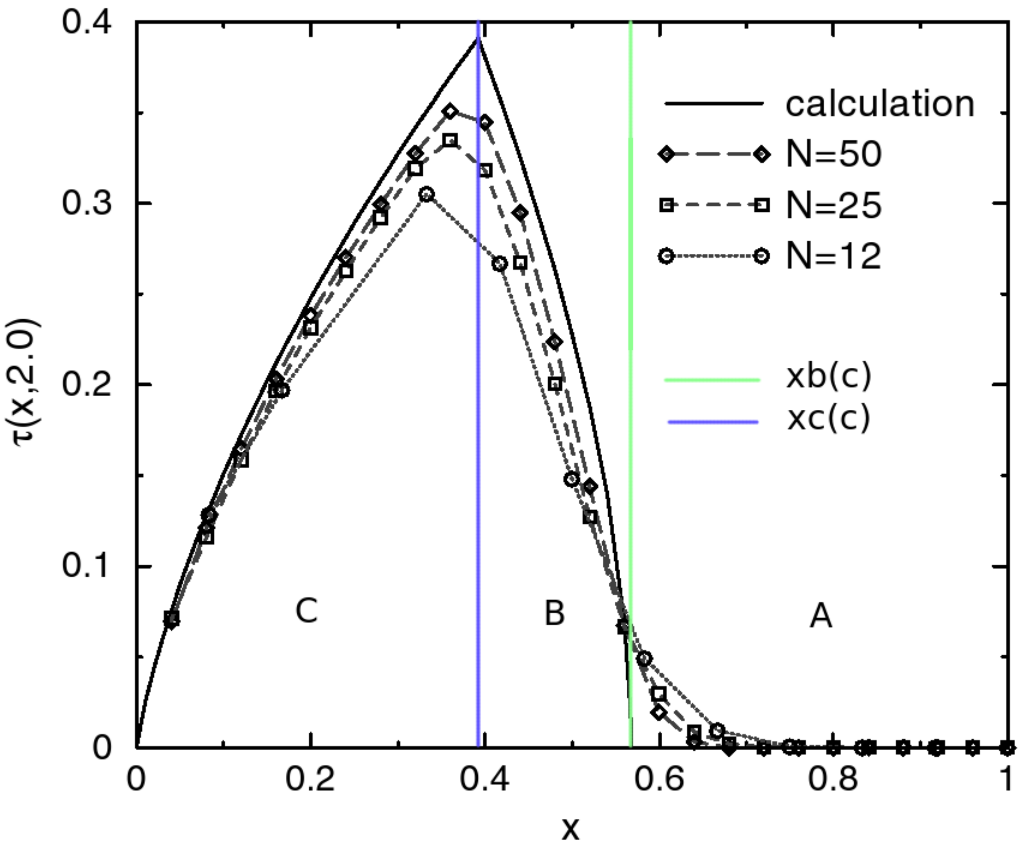


Figure 1: Experimental results for typical running time and the analytical solution. Graphs sampled from ensemble $\mathcal{G}(N, M)$. $\tau(x|c) = \lim_{N\to\infty} \frac{1}{N} \ln t_{bt}(\tilde{G}, \tilde{x})$ is the normalized and averaged logarithm of running time. C and B are the exponential phases and A is the linear phase. Dynamic phase boundary $x_b(c)$ and the static phase boundary $x_c(c)$ are also shown.

Original figure is from Martin Weigt and Alexander K. Hartmann, Typical Solution Time for a Vertex-Covering Algorithm on Finite-Connectivity Random Graphs, *Physical Review Letters*, **86**, 8 (2001).

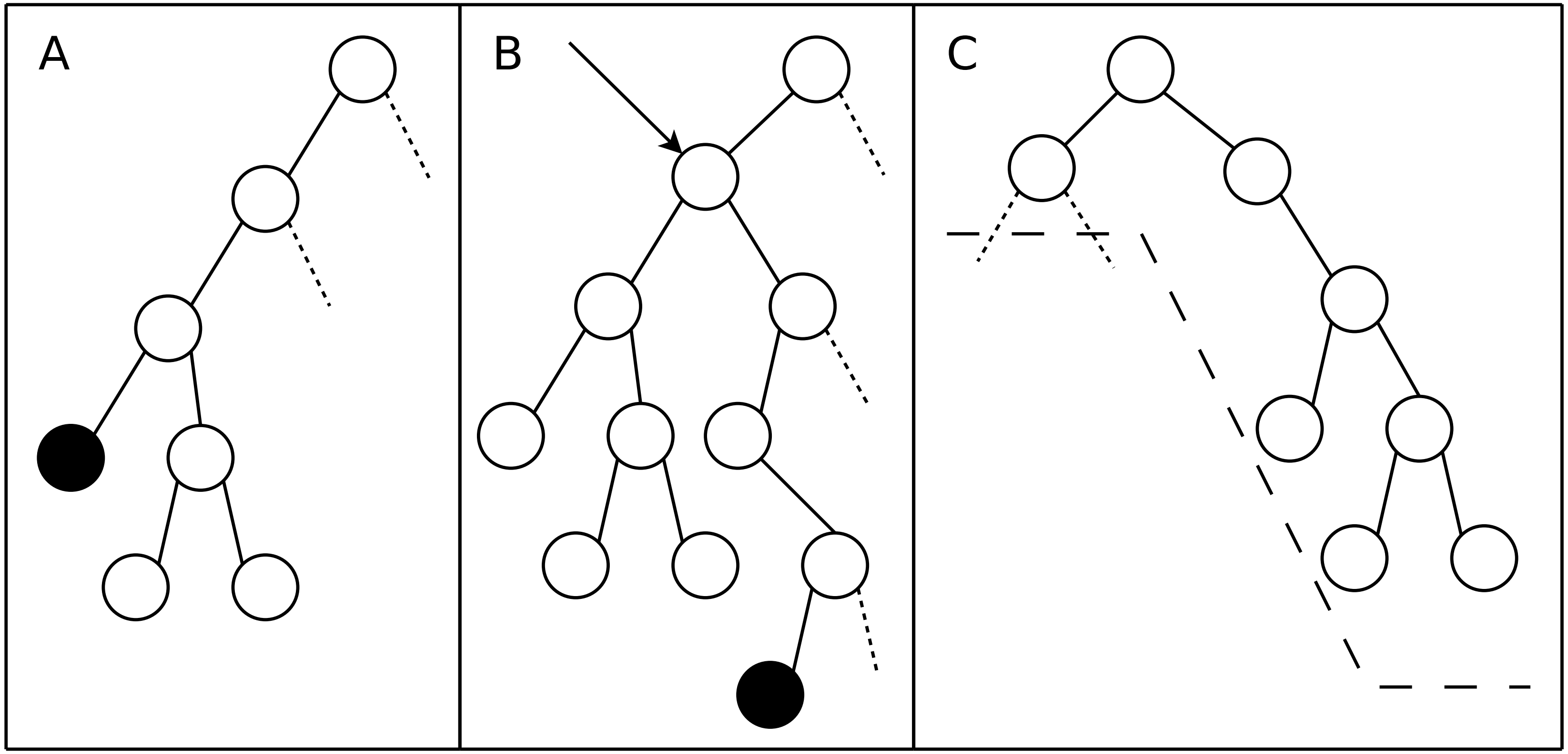


Figure 2: Example configuration trees. The three sub figures A, B and C show examples of configuration trees in the three different phases. In sub figure A the black circle is an example solution in the easy recoverable phase, where only a straight or almost straight descent is necessary. The black circle in sub figure B is an example solution in the hard recoverable phase, where we have to backtrack to the circle pointed to by the arrow. In sub figure C, in the hard unrecoverable phase, we have to traverse the tree until we know a solution can't be found (to the long dashed line).

First-moment method

- Goal: a lower bound on $x_c(c)$

- In the thermodynamic limit $N \to \infty$ for $x_1 < x_c(c) < x_2$, almost surely $|V_{vc}| \leq x_2 N$ exists, but $|V_{vc}| \leq x_1 N$ does not

- Use a fixed vertex and edge number ensemble $\mathcal{G}(N, \frac{c}{2}N)$

- Bound probability of a xN VC with the average number of VCs
  $G \in \mathcal{G}(N, \frac{c}{2}N)$
  $P(G \text{ has VC of size } xN) \leq \overline{\text{Number of VCs of size } xN}$
  $\Leftrightarrow P(\exists V_{vc}(G), |V_{vc}(G)| = xN) \leq \overline{\mathcal{N}_{vc}(G, xN)}$

First-moment method

- Number of potential vertex covers V_{vc}^* is $\begin{pmatrix} N \\ xN \end{pmatrix}$

- With probability $1-(1-x)^2 = x(2-x)$, V_{vc}^* covers at least one vertex of an edge

- All edges must be covered

$$\Rightarrow \overline{\mathcal{N}_{vc}(G,xN)} = \begin{pmatrix} N \\ xN \end{pmatrix} \underbrace{[x(2-x)]}_{P(\text{edge covered})}{}^{\overbrace{\frac{c}{2}N}^{\#\text{ edges}}}$$

First-moment method

- Approximation using Stirling's formula $\ln(N!) \simeq N \ln N - N$

$$\overline{\mathcal{N}_{vc}(G, xN)} = \left(\begin{array}{c} N \\ xN \end{array} \right) [x(2-x)]^{\frac{c}{2}N} = \frac{N![x(2-x)]^{\frac{c}{2}N}}{(xN)![(1-x)N]!}$$

$$= e^{\left\{\ln(N!)+\frac{c}{2}N \ln[x(2-x)]-\ln[(xN)!]-\ln[\{(1-x)N\}!]\right\}}$$

$$\simeq e^{\left\{N\left[-x\ln x-(1-x)\ln(1-x)+\frac{c}{2}\ln\{x(2-x)\}\right]\right\}}$$

First-moment method

- Exponent changes sign at $x_{an}(c) < x_c(c)$

  $0 = -x_{an}(c)\ln x_{an}(c) - (1 - x_{an}(c))\ln(1 - x_{an}(c)) + \frac{c}{2}\ln\{x_{an}(c)(2 - x_{an}(c))\}$

  $x_{an}(c) = 1 - 2\frac{\ln(c)}{c} + \mathcal{O}\left(\frac{\ln\{\ln(c)\}}{c}\right)$, for large average degree c

- Precise asymptotic:
  $x_c(c) = 1 - \frac{2}{c}(\ln(c) - \ln\{\ln(c)\} - \ln(2) + 1) + o(c^{-1})$

  A. M. Frieze, *Discr. Math.*, **81**, 171 (1990)

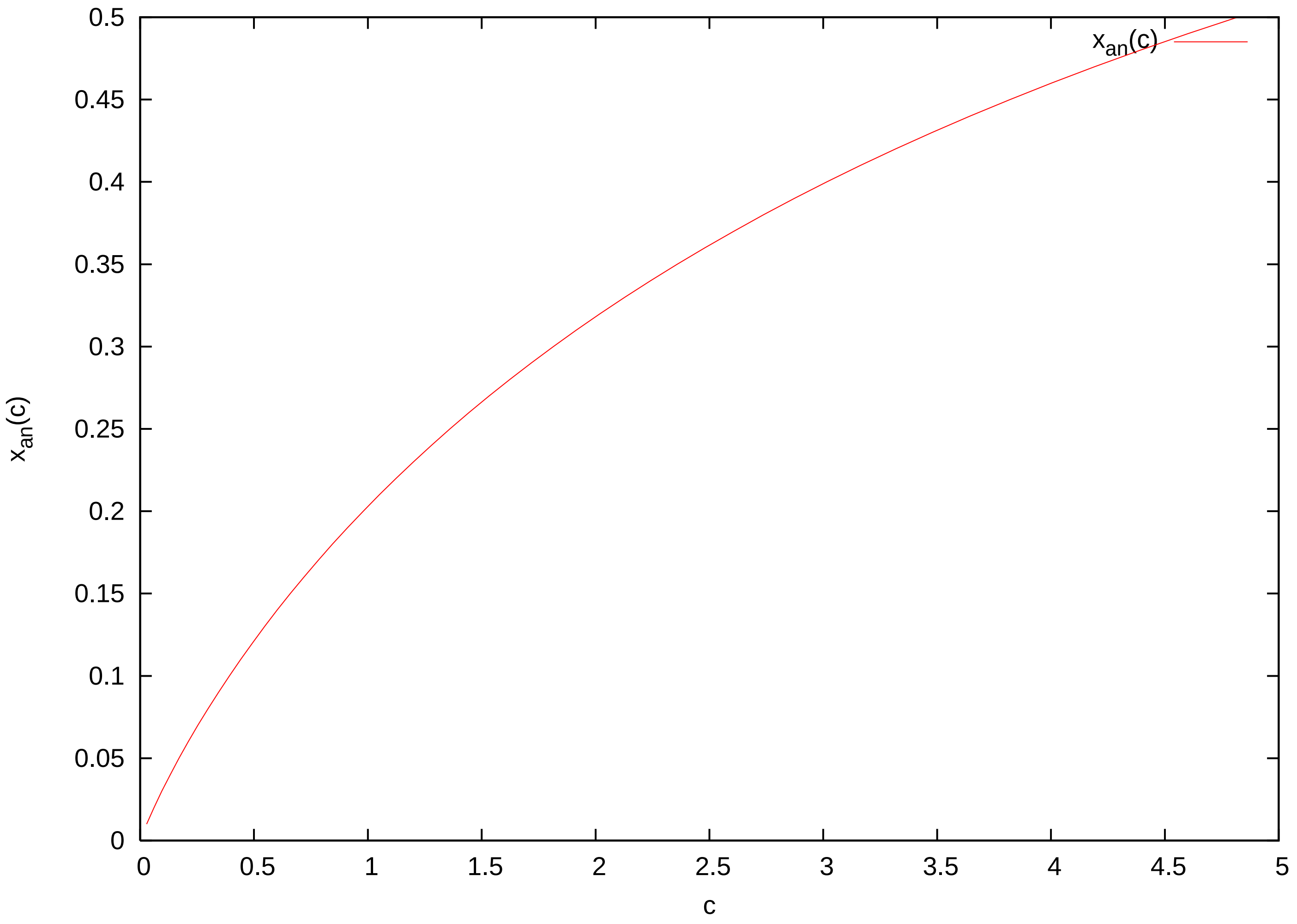


Figure 3: $x_{an}(c) \in [0, 0.5]$ plotted using formula $0 = -x_{an}(c)\ln x_{an}(c) - (1 - x_{an}(c))\ln(1 - x_{an}(c)) + \frac{c}{2}\ln\{x_{an}(c)(2 - x_{an}(c))\}$

Analysis of first descent into the configuration tree

- Consider covered vertices and edges as removed from graph
  - Then at time step T, $G \in \mathcal{G}(N-T, \frac{c}{N})$
  - Average edge degree $c(T) \simeq (N-T)\frac{c}{N} = (1-\frac{T}{N})c$
  - Rescaled time $t = \frac{T}{N}$, $G \in \mathcal{G}((1-t)N, \frac{c}{N})$

- For a single vertex at time t:

$$P(\text{vertex is not isolated}) = 1 - \underbrace{\left(1-\frac{c}{N}\right)}_{P(\text{no edge})}^{\overbrace{(1-t)N-1}^{\#\text{ possible edge endpoints}}}$$

$$\simeq 1 - e^{\{-(1-t)c\}}$$

Analysis of first descent into the configuration tree

- Available covering marks at time t:
  $X(t) = xN - N\int_0^t P(\text{vertex is not isolated})dt'$
  $= xN - N\int_0^t \left(1 - e^{\{-(1-t)c\}}\right) dt'$
  $= xN - Nt + N\frac{e^{-(1-t)c} - e^{-c}}{c}$

- First descent trajectory:
  $c(t) = (1-t)c$
  $x(t) = \frac{X(t)}{N(t)} = \frac{x-t}{1-t} + \frac{e^{-(1-t)c} - e^{-c}}{(1-t)c}$

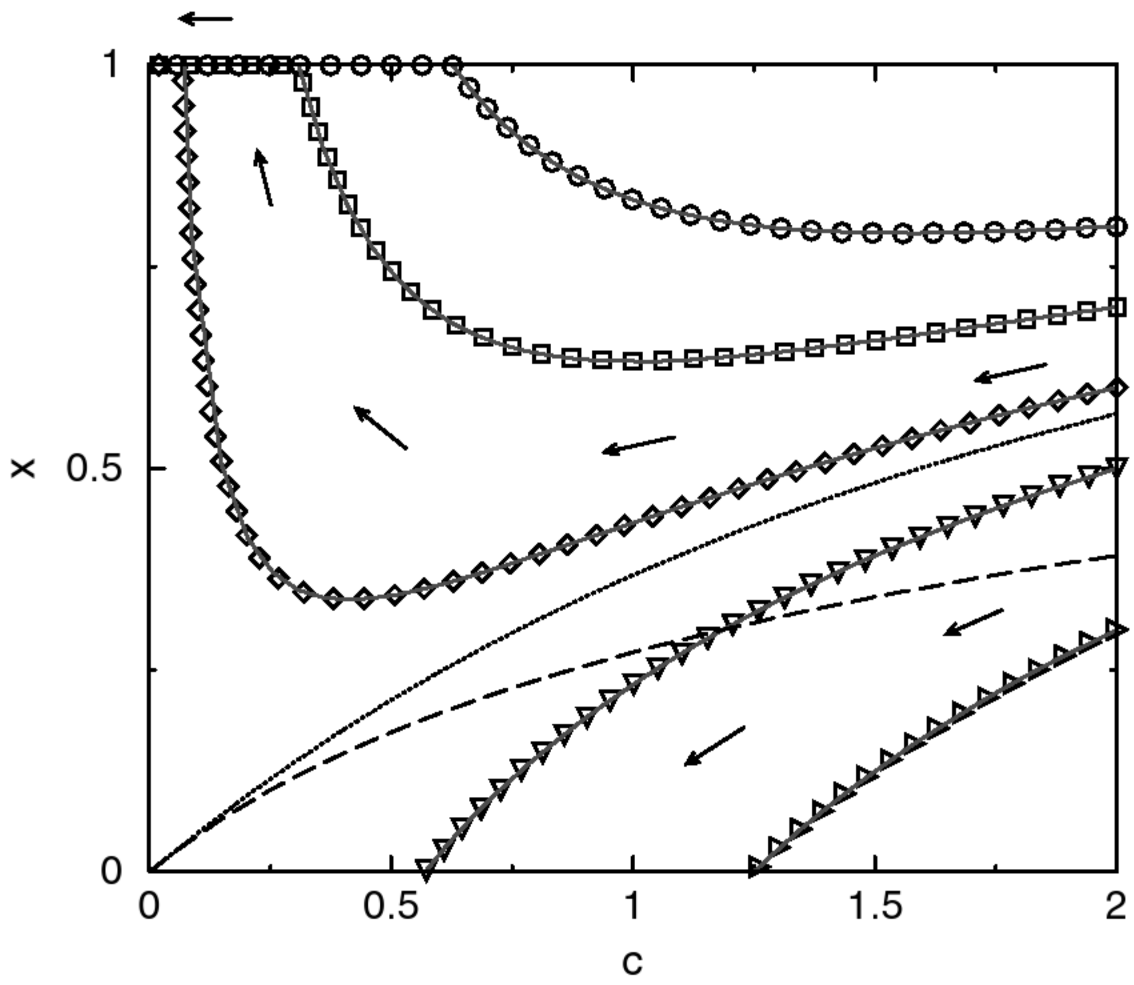


Figure 4: First descent into the configuration tree. Dotted line is $x_b(c)$ and long dashed line is $x_c(c)$. The lines start at $c = 2.0$ and at a certain x value.

Figure from Martin Weigt and Alexander K. Hartmann, Typical Solution Time for a Vertex-Covering Algorithm on Finite-Connectivity Random Graphs, *Physical Review Letters*, **86**, 8 (2001)

Analysis of first descent into the configuration tree

- Using $x(t' = 1) = 1$, we find the dynamical phase boundary $x_b(c)$:
  $x_b(c) = 1 + \frac{e^{-c}-1}{c}$

Backtracking

- Backtracking analysis can be difficult, because of configuration tree traversal dependencies

- Consider number of nodes at each configuration tree level

- At time $\tilde{t}$ the remaining covering marks $\tilde{x}\tilde{N}$ of the first descent are not enough

- $(\tilde{x}, \tilde{c})$ is the point, where $x_c(c)$ and the first descent cross (See figure 4)

- The uncoverable subtree $\tilde{G} \in \mathcal{G}(\tilde{N}, \frac{\tilde{c}}{\tilde{N}})$ must be backtracked in exponential time

Backtracking

- Exponential solution times log-normal distributed for large N
- $\Rightarrow$ Typical solution time $e^{N\tau(x,c)}$
  - backtracking time $t_{bt}\left(G_{\tilde{N},\frac{\tilde{c}}{\tilde{N}}},\tilde{x}\right)$ for uncoverable $G_{\tilde{N},\frac{\tilde{c}}{\tilde{N}}}$
  - quenched average $\tau(x,c)=\lim_{N\to\infty}\frac{1}{N}\overline{ln\left[t_{bt}\left(G_{\tilde{N},\frac{\tilde{c}}{\tilde{N}}},\tilde{x}\right)\right]}$

⋮

- Final result for the bounded algorithm:

  $\tau(x,c)\simeq\max_{\kappa=\tilde{x},\ldots,1}\left[\frac{c}{\tilde{c}}\kappa s_{an}\left(\frac{\tilde{x}}{\kappa},\tilde{c}\kappa\right)\right]$

  $s_{an}(x^*,c^*)=-x^*lnx^*-(1-x^*)ln(1-x^*)+\frac{c^*}{2}ln(x^*\left[2-x^*\right])$

  (See figure 1)

Summary

- Three different phases and two boundaries $x_b(c)$ and $x_c(c)$ observable in experiments

- First-moment method
  - A lower bound on $x_c(c)$
  - Using the average vertex cover amount as a bound

- First descent analysis shows where exponential solution times start occurring: dynamical phase boundary $x_b(c)$

- Backtracking analysis: calculate time to backtrack the largest subtree, get subtree root using first descent results and $x_c(c)$