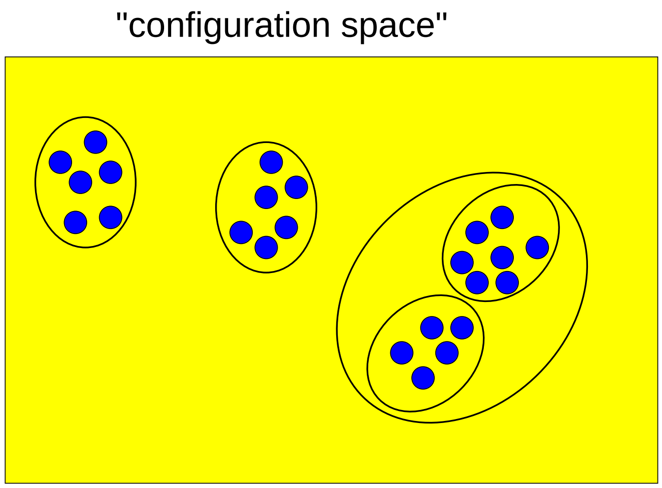
Phase transitions in combinatorial optimization problems
Course at Helsinki Technical University, Finland, autumn 2007
by Alexander K. Hartmann (University of Oldenburg)
Lecture 7, 9. October 2007

3.8 Clustering

Usually: min. VCs not distributed uniformly in conifg space.

Clusters := (sloppy definition) groups of min. VCs that are separated by regions where no min. VCs exist.



Possible: Hierarchy of clustering

Important question: clustering related to computational hardness?

Physics: infinite hierarchy observed for spin glasses (SK model, defined on a complete, i. e., fully connected graph)
Analytically: corresponds to replica-symmetry breaking (RSB)
Ising ferromagnet: no clustering
Most models: no analytical clustering possible.
→ study clustering using numerical methods. Here: VC

3.8.1 Neighbor-based clustering

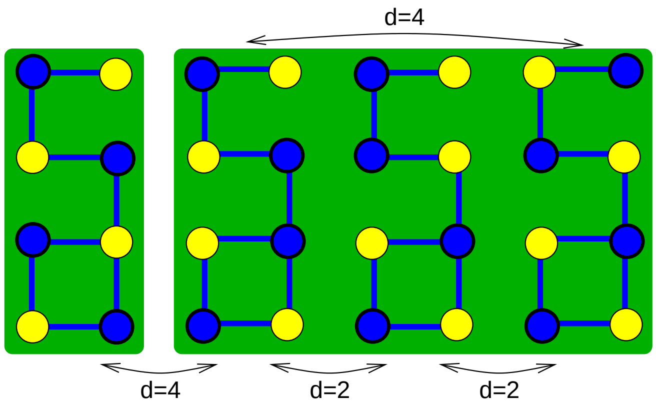
Given: set of configurations $\{\underline{c}^\alpha\}$ ($c_i^\alpha = 1$ if $i \in$ VC α, 0 else)

Hamming distance

$$d_{\text{Hamming}}(\underline{c}^\alpha, \underline{c}^\beta) = \sum_i |c_i^\alpha - c_i^\beta| \tag{1}$$

$\underline{c}^\alpha, \underline{c}^\beta$ "neigbors" ⇔ $d_{\text{Hamming}}(\underline{c}^\alpha, \underline{c}^\beta) \leq d_{\max}$. (example: VC ($d_{\max} = 2$)).
Cluster: transitive closure of neighbour relation.

Example: Clusters for VC

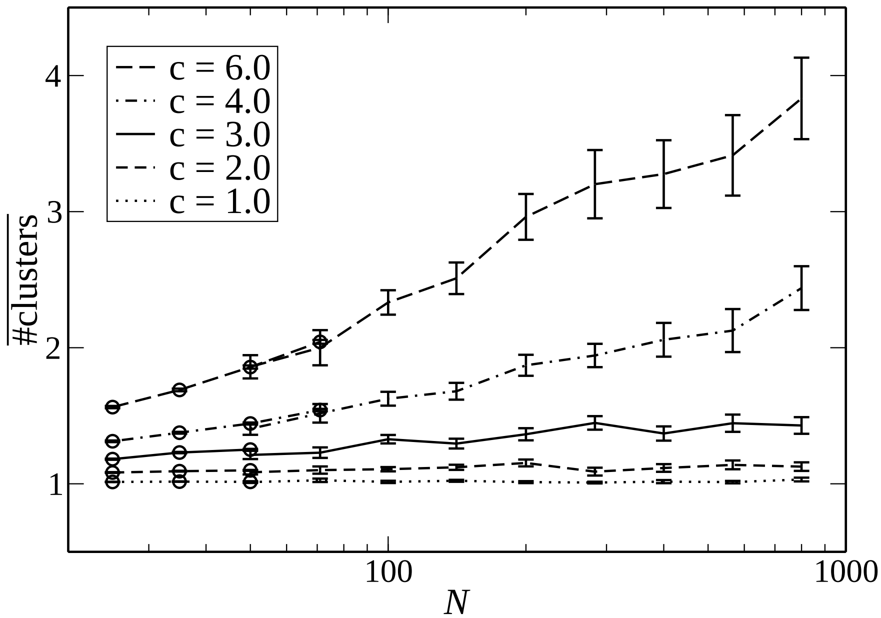


→ two clusters □

Algorithm: grow clusters by adding neighbors, $O(\#\text{VCs}^2)$
Assumption: Set S of all VCs available

begin
    $i = 0$ {number of so far detected clusters}
    while S not empty **do**
    begin
        $i = i + 1$
        remove an element $\underline{c}^{(\alpha)}$ from S
        set cluster $K_i = \{\underline{c}^{(\alpha)}\}$
        set pointer p to first element of K_i
        while $p <> NULL$ **do**
        begin
            for all elements $\underline{c}^{(\gamma)}$ of S
                if $d_{ham}(p, \underline{c}^{(\gamma)}) \leq d_{\max}$ **then**
                begin
                    remove $\underline{c}^{(\gamma)}$ from S
                    put $\underline{c}^{(\gamma)}$ at the end of K_i
                end
            set pointer p to next element of K_i
            or to $NULL$ if there is no more
        end
    end
end

Result:



$c < e$:	ONE cluster, independent of N
$c > e$:	several clusters, logarithmic growth in N

corresponds to onset of RSB in analytical solution

Note: large systems: too many solutions
→ generate sample using parallel tempering
then use "ballistic search" for clustering (see W. Barthel and A.K. Hartmann, "Clustering analysis of the ground-state structure of the vertex-cover problem", Phys. Rev. E **70**, 066120 (2004)).

or use:

3.8.2 Hierarchical clustering

Aim: represent cluster structure as tree.
Input:

- Sample set of "items", e.g. configurations $\{c^\alpha\}$ sampled in equilibrium, or min. VCs.
- Distances $d(\underline{c}^\alpha, \underline{c}^\beta)$

Initially:

- Each item → one cluster
  $K_\alpha := \{\underline{c}^\alpha\}$ with size $n_\alpha = 1$
- Set of clusters $S := \{K_\alpha\}$
- Cluster distances ("proximiy matrix")
  $d_{\alpha,\beta} = d(\underline{c}^\alpha, \underline{c}^\beta)$

Algorithm ("agglomerative clustering")

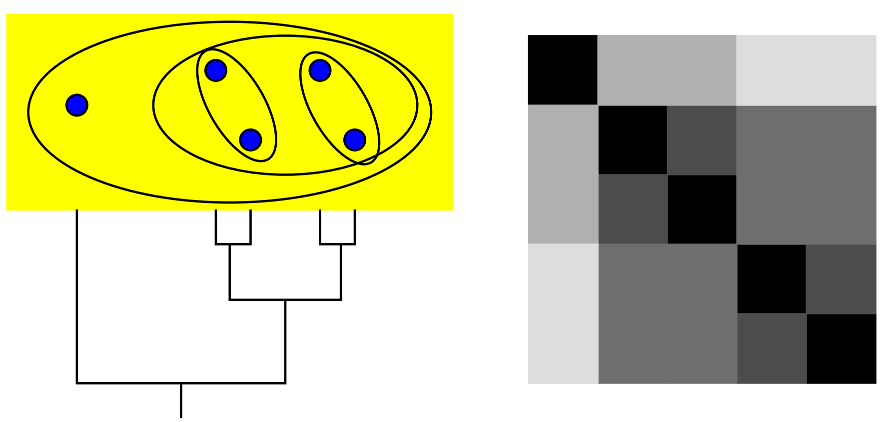
algorithm
begin
while there is more than one cluster **do**
  begin
    Select two clusters K_α, K_β with the minimal distance
    Merge Clusters $K_\gamma := K_\alpha \cup K_\beta$
    for all other clusters K_δ **do**
      update $d_{\gamma,\delta} = \ldots$
  end
end

Tree representation ("dendrogram"):

- Single items = leaves
- Merge of clusters = two subtrees (daughters) meet in mother node
- Length of edge of node to mother = distance of its two daughters when merged

.
Order of leaves (not unique) → ordering of items
→ Proximity matrix: grey shaded drawn with rows/colums in that order
→ cluster structure becomes visible.

Example: General idea



□

Different choices for update function possible
Here: Ward's method

$$d_{\gamma,\delta} = \frac{(n_\alpha + n_\delta)d_{\alpha,\delta} + (n_\beta + n_\delta)d_{\beta,\delta} - (n_\delta)d_{\alpha,\beta}}{n_\alpha + n_\beta + n_\delta} \qquad (2)$$

<u>Example</u>: 4 configuration from Sec. 3.8.1
$S = \{K_1, K_2, K_3, K_4\}$, $n_1 = 1, n_2 = 1, n_3 = 1, n_4 = 1$

$$(d_{\gamma,\delta}) = \begin{pmatrix} 0 & 4 & 6 & 8 \\ 4 & 0 & 2 & 4 \\ 6 & 2 & 0 & 2 \\ 8 & 4 & 2 & 0 \end{pmatrix}$$

<u>Iteration 1</u>: $K_{2'} = K_2 \cup K_3$, $n_{2'} = 2$
$d_{2'1} = \frac{(1+1)4+(1+1)6-1\cdot 2}{1+1+1} = \frac{18}{3} = 6$
$d_{2'4} = \frac{(1+1)4+(1+1)2-1\cdot 2}{1+1+1} = \frac{10}{3} = 3.333$

$$\Rightarrow \quad (d_{\gamma,\delta}) = \begin{pmatrix} 0 & 6 & 8 \\ 6 & 0 & 10/3 \\ 8 & 10/3 & 0 \end{pmatrix}$$

<u>Iteration 2</u>: $K_{4'} = K_{2'} \cup K_4$, $n_{4'} = 3$
$d_{4'1} = \frac{(2+1)6+(1+1)8-1\cdot 10/3}{2+1+1} = \frac{102/3-10/3}{4} = \frac{92}{12} = \frac{23}{3}$

$$\Rightarrow \quad (d_{\gamma,\delta}) = \begin{pmatrix} 0 & 23/3 \\ 23/3 & 0 \end{pmatrix}$$

<u>Iteration 3</u>: $K_{1'} = K_1 \cup K_{4'}$, $n_{1'} = 4$

□

Results for vertex cover:

(small μ):	no structure ("paramagnet")
$c < e$:	solution cluster has no structure
$c > e$:	hierarchy of solution clusters