Sequential Superparamagnetic Clustering

-

a Network Self-organisation Process

PASCAL January 2005

Thomas Ott, Albert Kern and Ruedi Stoop
Institute of Neuroinformatics, ETH Zürich

I would like to point out the following points:

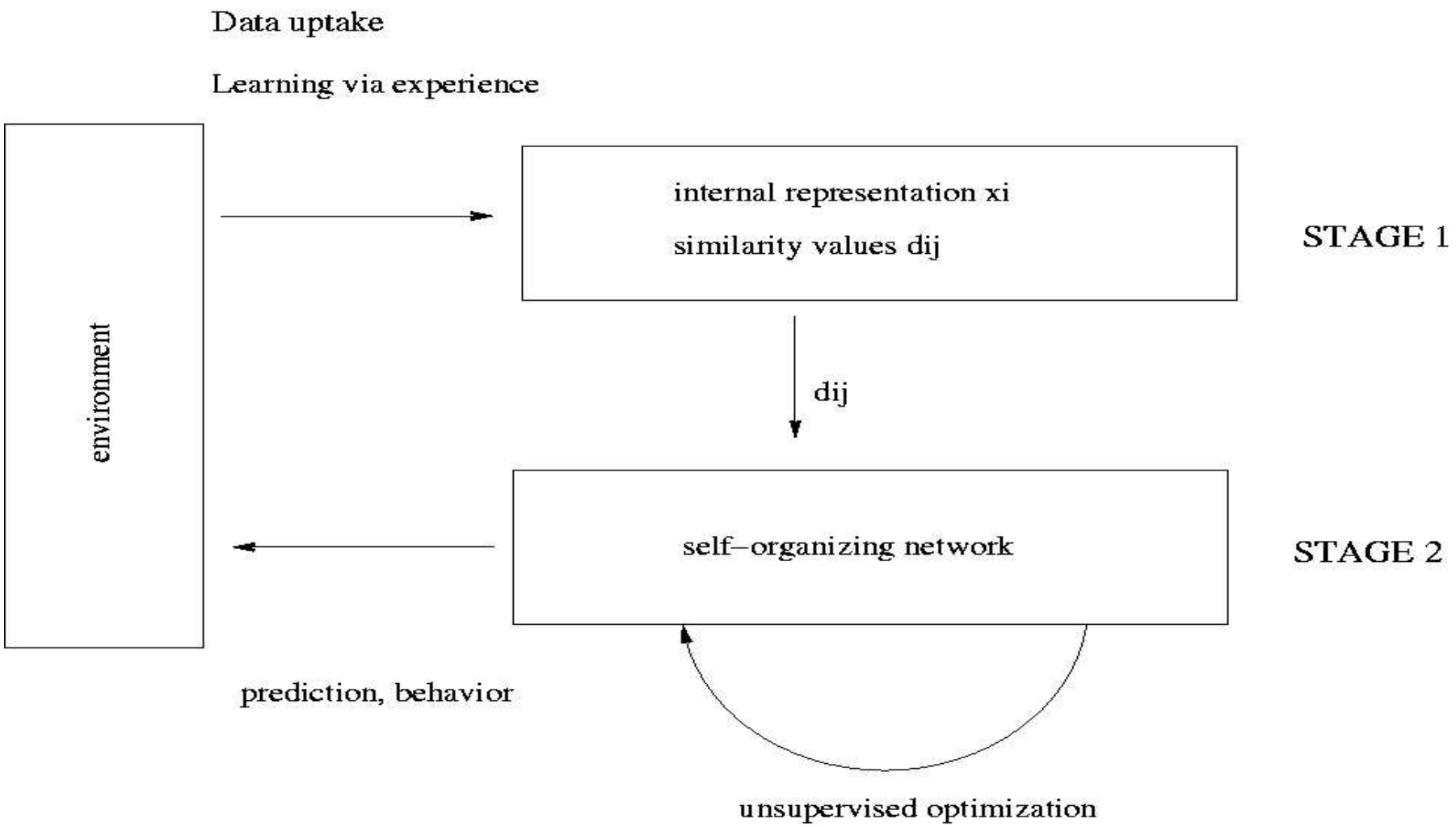
- Clustering is a useful tool for unsupervised classification talks (e.g., for scene analysis) => cognition-like operation.
- We need unbiased methods.
- Superparamagnetic clustering comes close to an 'ideal' method.
    - Short introduction.
    - Example.
    - Example of failure.

- Idea of how to remedy the failure.
- Sequential superparamagnetic algorithm.
    - Example.

Clustering: Grouping together items / points / parts ...
that belong together.

- How to represent the single objects ?
- Best similarity measure ?
- Number of classes normally unknown.
- Inherent branching nature: Best resolution ?

Sequential Superparamagnetic Clustering – a Network Self-organisation process



Superparamagnetic Clustering: (Blatt, Wiseman, Domany 1996)

- n data items <=> n Pott spins $s_i \in \{1,...,q\}$ with q = typically 10 or 20.
  (<=> n neurons with q firing states).

- Connectivity: Each neuron coupled to its k nearest neighbours (not necessarily mutual neighbours) with

$$J_{ij} = 1/K \; e^{-d_{ij}^2 / 2a^2} .$$

d_{ij}: dissimilarity between item i and j.
K: average number of coupled neighbours.
a: average dissimilarity of coupled neighbours.

Superparamagnetic Clustering: (Blatt, Wiseman, Domany 1996)

- Hamiltonian: $H(s) = \sum_{(i,j)} J_{ij}(1 - \delta_{s_i s_j})$.

- Boltzmann/Gibbs: $P(\mathbf{s}, T) = 1/Z(T)\ \mathrm{Exp}(-H(\mathbf{s})/T)$.

  (T: Temperature/ noise control parameter).

- Cluster detection: If $G_{ij} = \sum_s p(s,T)\ \delta_{si\ sj} > \Theta$

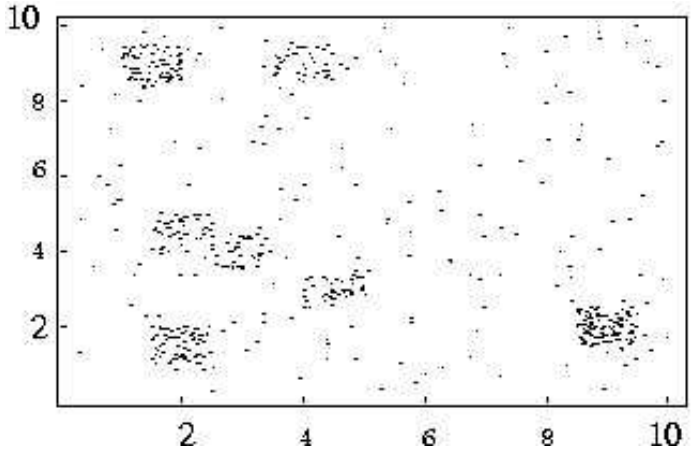
  then i and j belong to the same cluster.

  (transitive: $(i,j) \in C \ \wedge\ (j,k) \in C \Rightarrow (i,k) \in C$.)

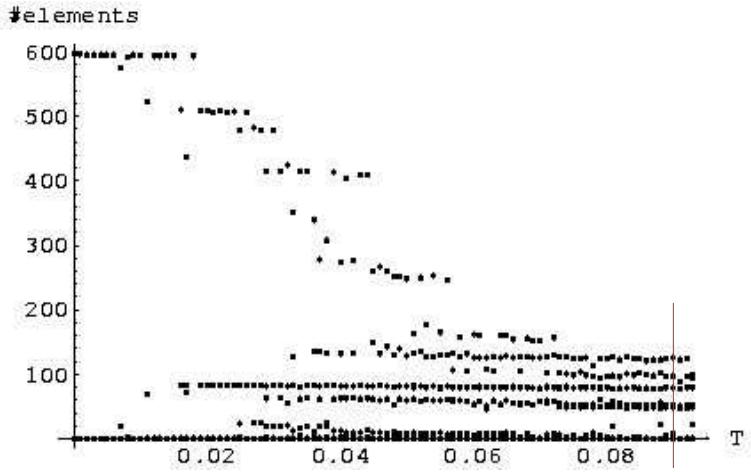
- Update: Any appropriate MCMC
  (e.g., Swendsen-Wang algorithm).

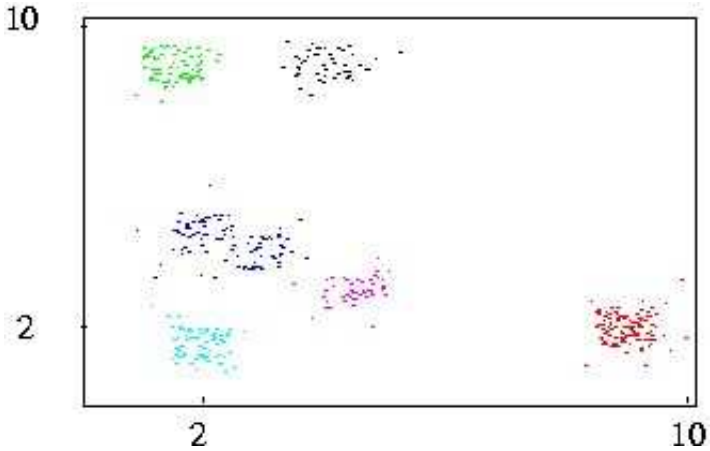
Superparamagnetic Clustering **Example**:

Temperature T controls the resolution => hierarchy !

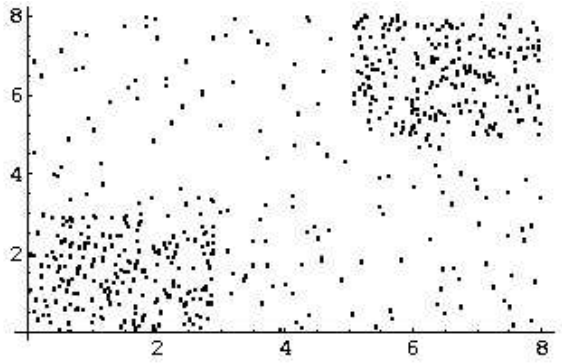


knN = 10,
q = 20 .

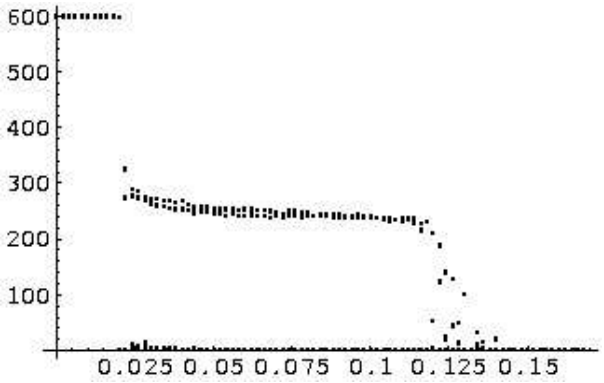




T = 0.09



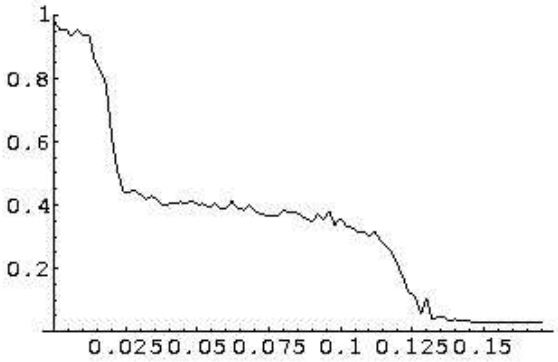
Cluster-size



T

$$M = \frac{(\frac{N_{max}}{N} q - 1)}{(q-1)}$$

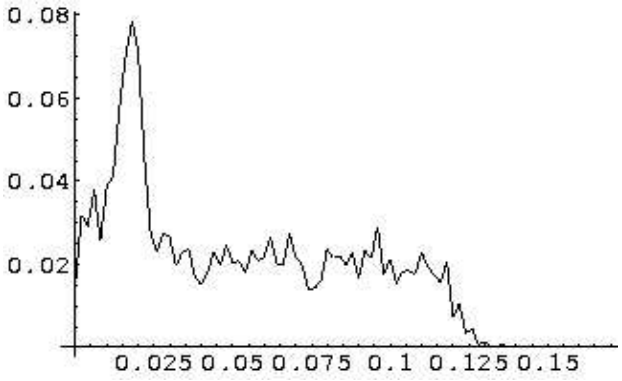
<M>



T

$N_{max} = \max \{N_1, \ldots N_q\}$

χ *T/N



T

χ *T/N = $(\langle M^2 \rangle - \langle M \rangle^2)$

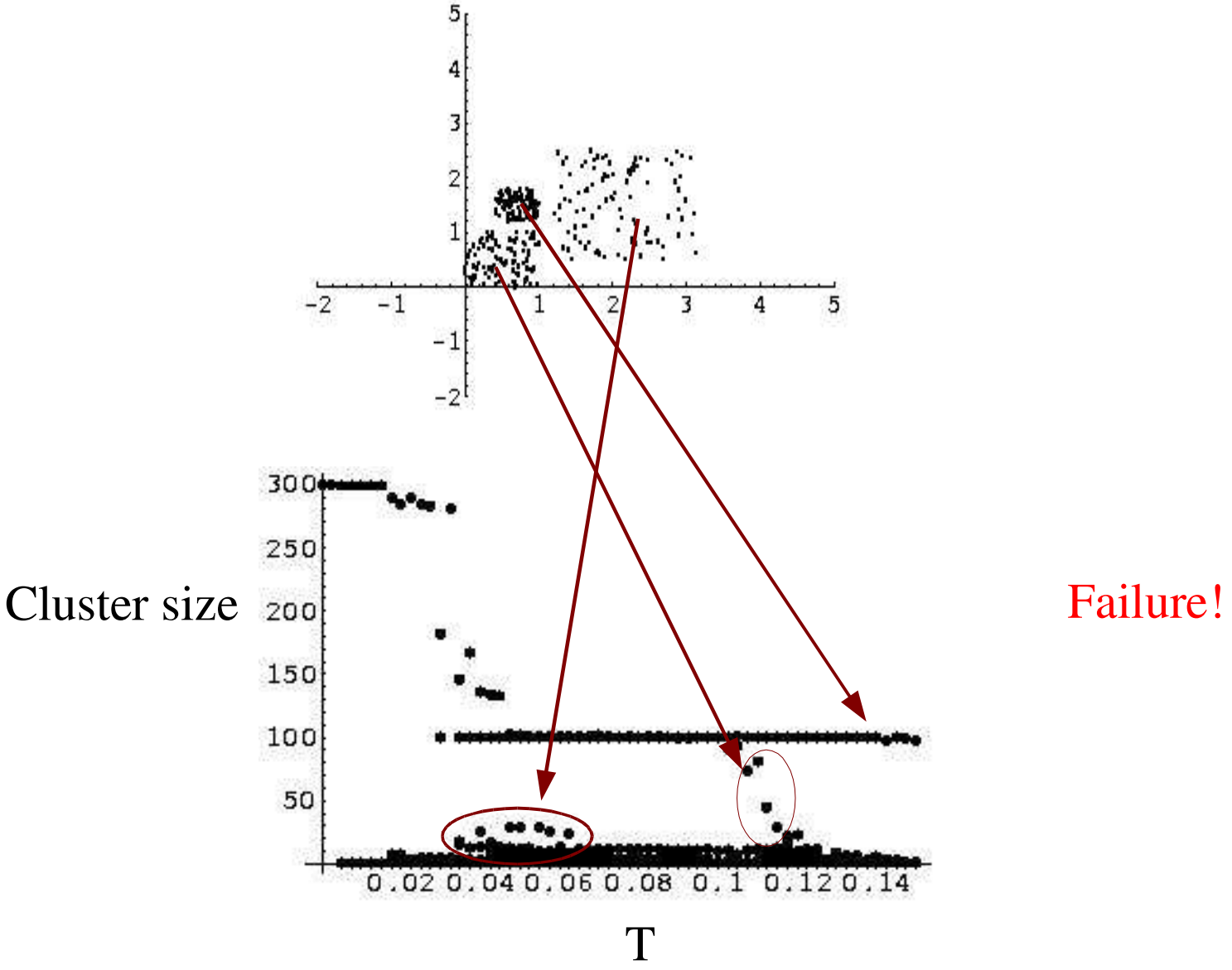
Superparamagnetic Clustering:

Positive:
- No bias for shape or number of clusters.
- Allows for extracting different clusters at different resolution levels.
- Robustness: $G_{ij} \approx 0$ or $G_{ij} \approx 1 \Rightarrow 1/q < \Theta < 1 - 2/q$.
- T-stability is a good estimator for the compactness of a cluster.

Negative:
- How to extract the most natural clusters ? (what are criterion for 'natural'?)
- How to choose k (k nearest neighbours) ?

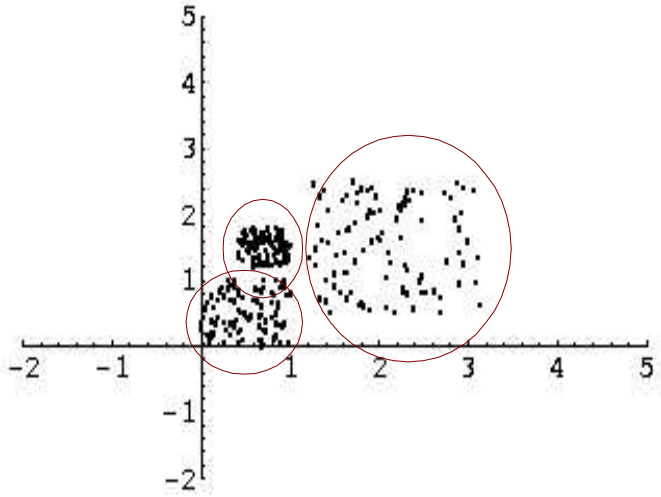
- next Example:
  => How to deal with clusters of different densities ?

Superparamagnetic Clustering:



Failure!

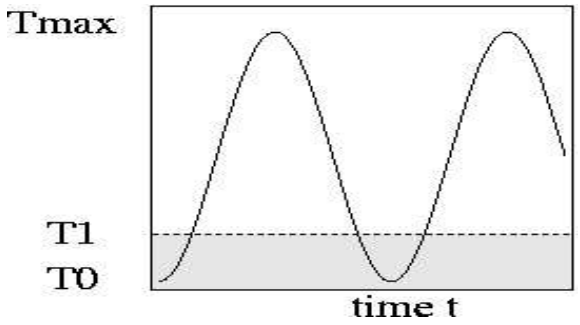
***Ultimate* question:** Can the difficulties be overcome ?



Idea: Find learning rules for the network structure that
1) decouple the 'obvious' clusters,
2) locally reset the 'synaptic weights' (by readjusting K and a).

Gain: Substructures are found on a more local, i.e. mesoscopic, scale.
I.e., clusters are found on different, but appropriate resolution levels (cascade-like decoupling).

=> Hebbian-like learning rule (semi-local) : concept



} Period of readjustment J_{ij}^{max}

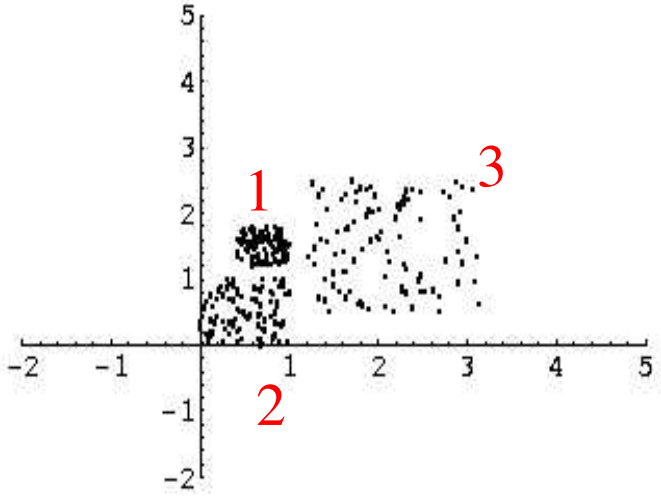
$$\Delta_T J_{ij} \sim -c\, J_{ij}^{max} \left(1 - G_{ij}(T)\right) f\left(g_i(T), g_j(T)\right)$$

$$g_i = \langle G_{ik} \rangle_k \,,\; g_j = \langle G_{jk} \rangle_k$$

$$f(g_i \approx 1, g_j \approx 1) = f(g_i \approx 0, g_j \approx 1) = f(g_i \approx 1, g_j \approx 0) = 1$$

$$f(g_i \approx 0, g_j \approx 0) = 0$$

$$\Delta_T J_{ij} \sim -c\, J_{ij}^{max} \left(1 - G_{ij}(T)\right) f\left(g_i(T), g_j(T)\right)$$



1) Connections within clusters 1,2,3 are not weakened.

Either $f(\approx 1, \approx 1) = 1$ and $G_{ij}=1$ or $f(\approx 0, \approx 0) = 1$ and $G_{ij}= 0$.

2) Connections within different clusters are weakened.

$f(\approx 0, \approx 1) = 1$ and $G_{ij}= 0$ for a certain T-range .

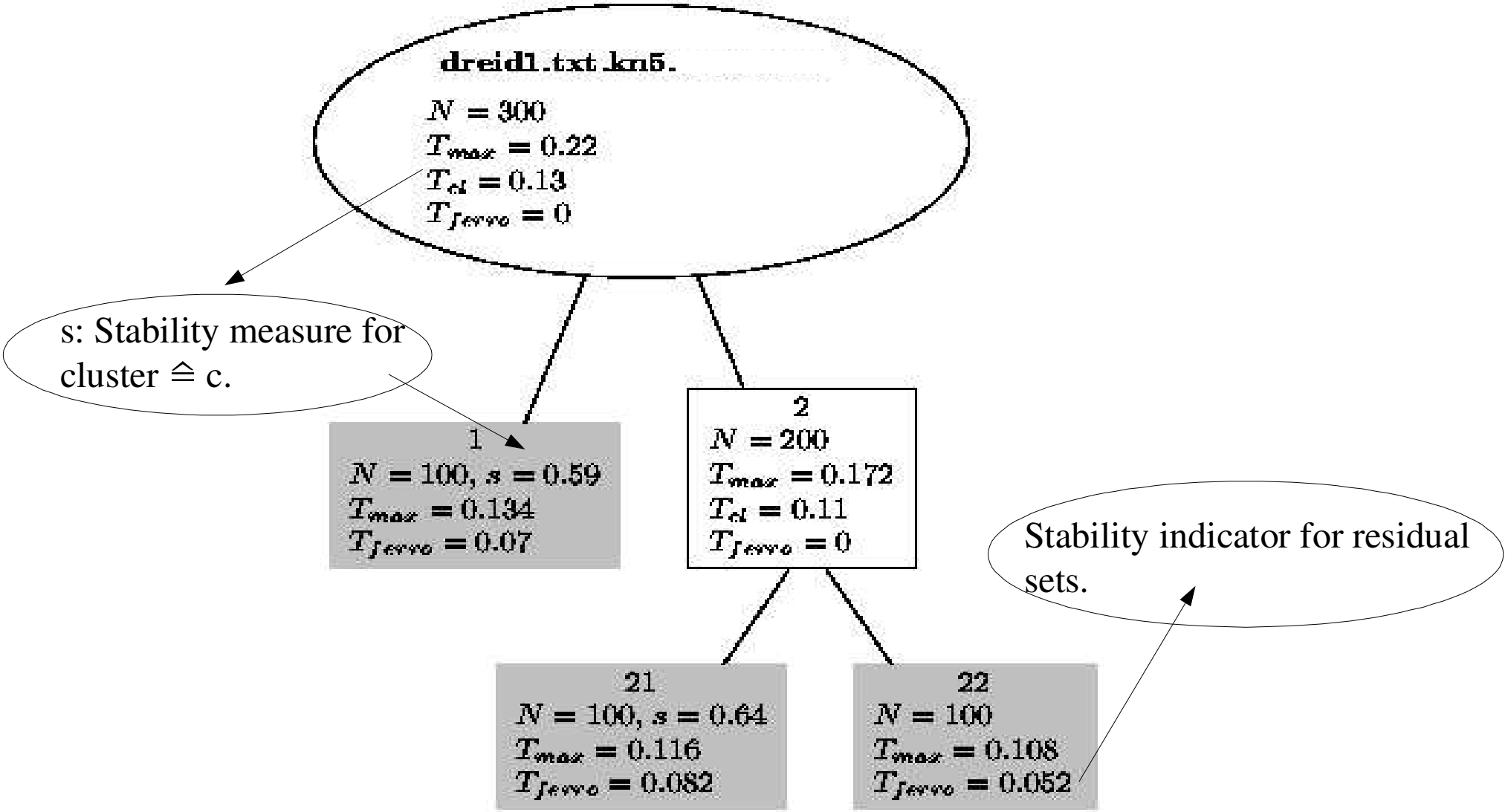
3) c regulates how stable (T-range) a cluster must be to be decoupled.

Effective implemented algorithm:

Sequential superparamagnetic clustering SSC:

Extract the most stable cluster, i.e. the cluster that is stable over the longest T-range, and cluster this set and the residual set separately until no stable subclusters can be found any more

=> off-line variant.



Dendrogram as result from the SSC procedure.

Sequential superparamagnetic clustering SSC:

1) If stability s < threshold, e.g., 0,2 then end.

=> one global *parameter of interest* (cut-off),
but locally regulated resolution levels.

2) Residual set: data points that has not been extracted, but don't have a substructure.
T_{ferro} indicates whether the set can be interpreted as natural cluster or just as a background distribution.

Successful application of SSC:

T.Ott,A.Kern,A.Schuffenhauer,M.Popov,P.Acklin,E.Jacoby and R.Stoop
Sequential Superparamagnetic Clustering for Unbiased Classification of High-dimensional Data,
J.Chem.Inf.Comp.Sci. (July 2004).

Thank you