M.C. Thrun, AG Datenbionik bei Prof. Ultsch

Introduction in the Concept of Swarms

Chapter 7: Behavior-Based Systems in Data Science, Dissertation

Philipps Universität Marburg

Anzahl Schwarmsysteme in DataMining

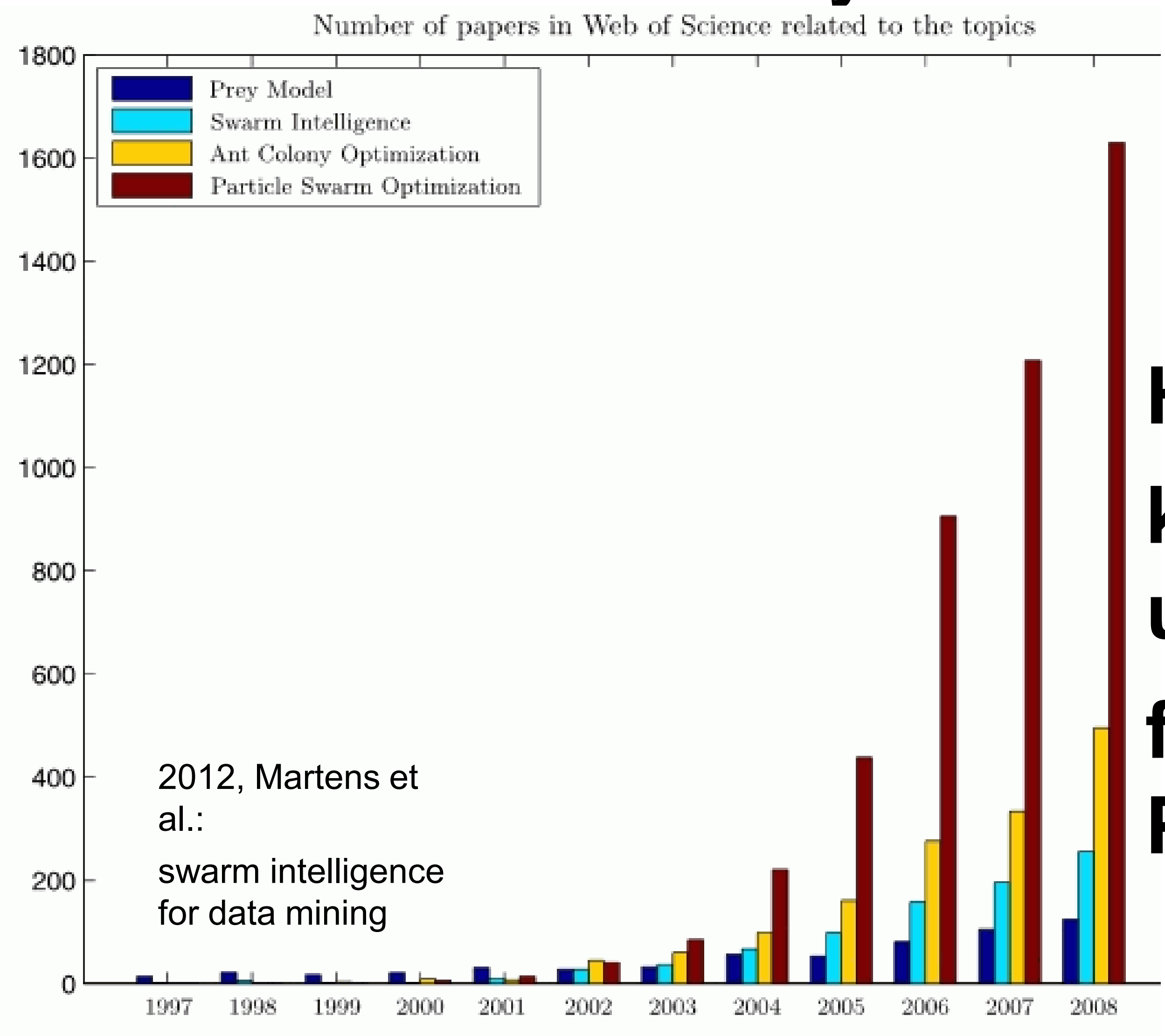
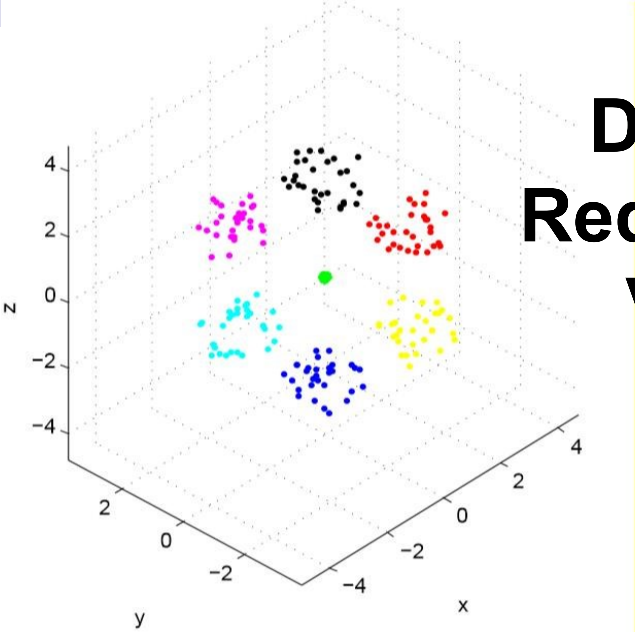
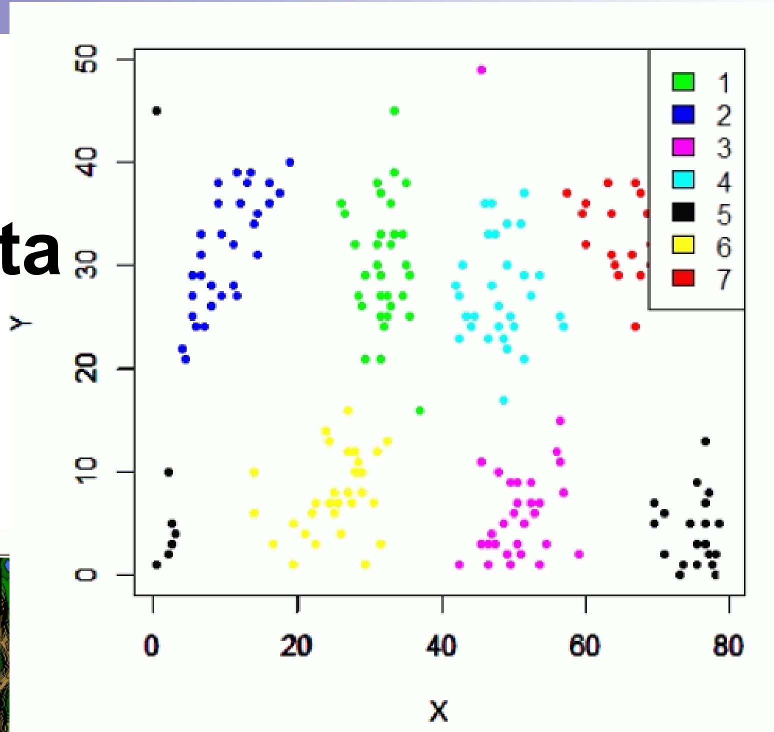


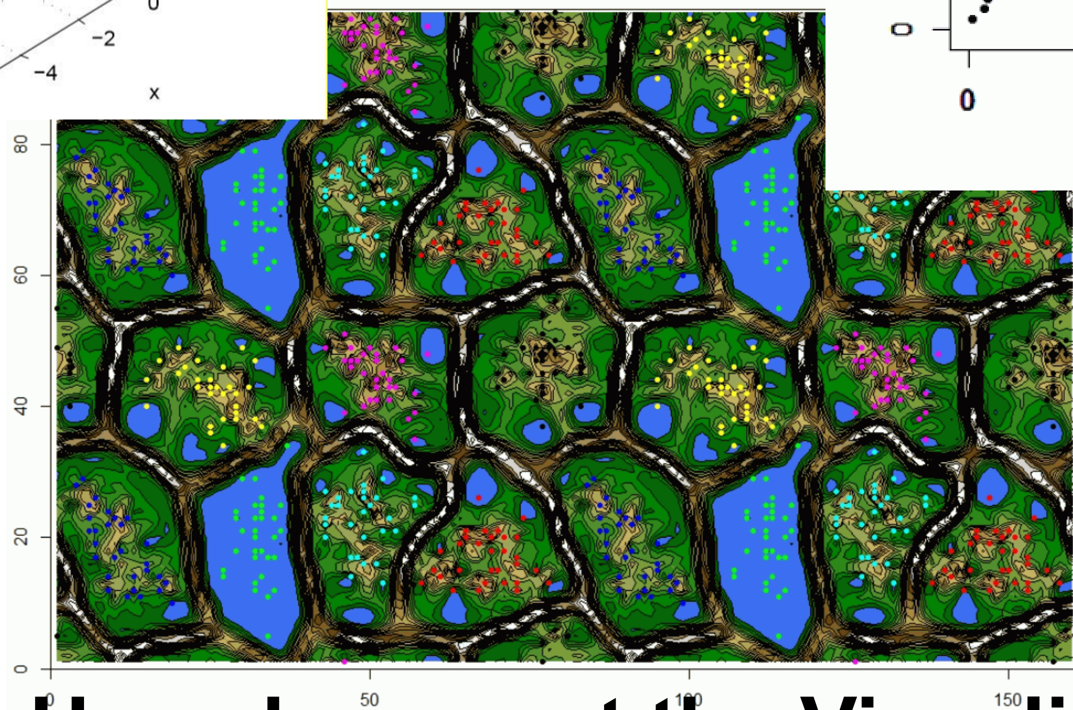
Fig. 1 Number of papers written on the subject, from Web of Science

Hier: nur ein kleiner Auschnitt unter folgender Problemstellung



Dimensionality Reduction for Data Visualization





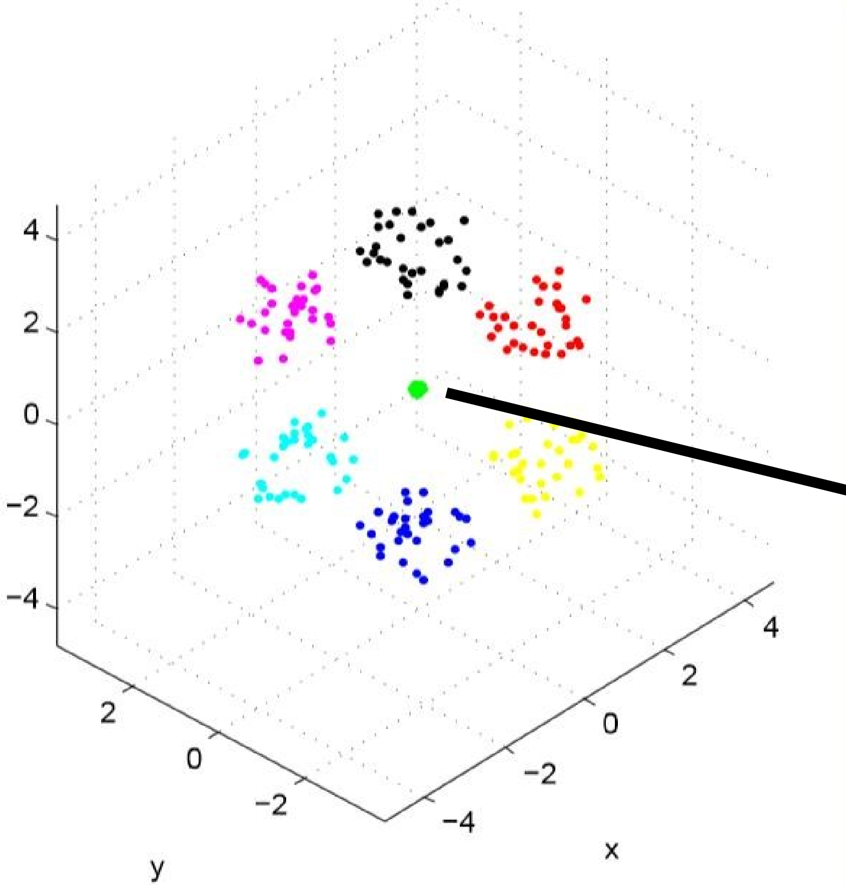
How do we get the Vizualization?

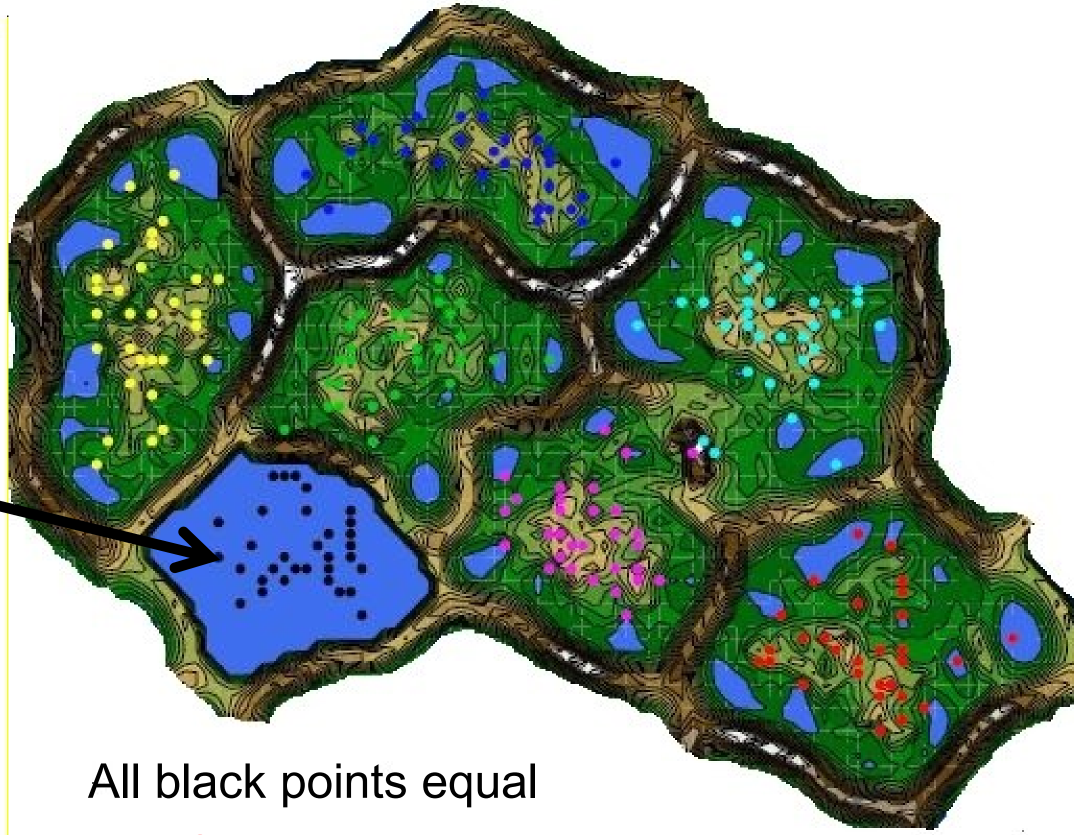
Problem

Problem: separate data into similar groups -> **Clustering**

Solution: Visualize data in two dimensions as landscape -> **Projection**

Goal: detect and visualize meaningful cluster structures





All black points equal one Cluster

Concepts for Swarms

- Swarm intelligence [Beni 1989]
- Self-Organization
  - Swarms: [Bonabeau/Dorigo et al., 1999]
  - SOM: [Ultsch 1992]
- Bionics [Deneubourg 1991, Reynolds, 1987]

"the application of biological methods and systems found in nature"

- **Missing links** [Thrun, 2016]:
  - Application of Game Theory
  - Emergence

What is Intelligence?

- In general the definition of intelligence is controversial [Legg/Hutter, 2007] and complex [Zhong, 2010].

- In the context of swarms, behavior and intelligence are used synonymously

Swarm Intelligence (SI)

- **the emergent collective behavior of simple entities called agents [Bonabeau et al., 1999, p. 12], [Martens et al., 2011, p. 2]**

- **Collective behavior generically denotes any behavior of agents in a system having more than one agent [Cao et al., 1997]**

- **An agent is a software entity**
  - **situated in an environment**
    - "situated," (…) is intended to emphasize that the process of deliberation takes place in an agent that is directly connected to an environment" [Russell et al., 2003, p. 422]
    - the agent receives sensory input from its environment and that it can perform actions which change the environment in some way
  - **capable of flexible, autonomous action in order to meet its design objectives [Jennings et al., 1998]**

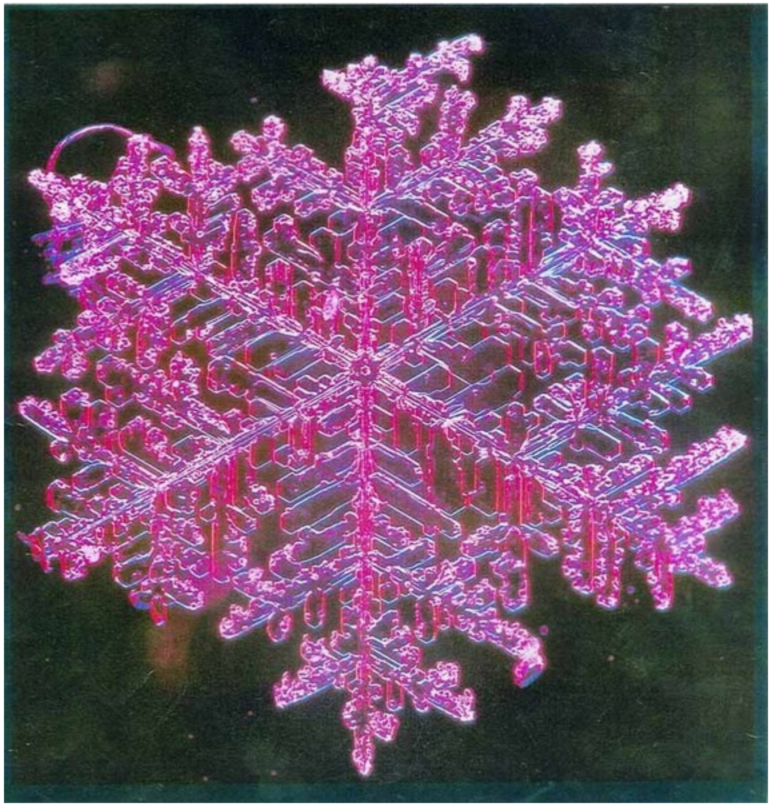
Five principles of swarm behavior

1. *Homogeneity,*
   1. every agent has the same behavior model;
2. *Locality*
   1. the motion of each agent is only influenced by its nearest neighbors;
3. *Velocity matching*
   1. every agent attempts to match the velocity of nearby flock mates
4. *Collision Avoidance*
   1. every agent avoids collisions with nearby agents
5. *Flock Centering*
   1. the agents attempt to stay close to the neighboring agents

Self-organization (SO)

Self-organization is defined by spontaneous pattern formation by a system itself, without responsibility of any determinate inside agent.

Fish swarm



Ice flake

Four basic ingredients for SO in a swarm

1. *positive feedback*
   - promote a creation of convenient structures and help to stabilize them
2. *negative feedback*
   - promote a creation of convenient structures and help to stabilize them
3. *Amplification of fluctuations*
   - *fluctuations* defined as errors, random movements and task switching.
4. *multiple interactions*.
   1. For swarm behavior to emerge, multiple interactions are required.

[Grosan et al., 2006]

Emergence

- An ability of an system
- the arising of novel and coherent properties during the process of self-organization [Goldstein 1999, Ultsch 1999, 2007].

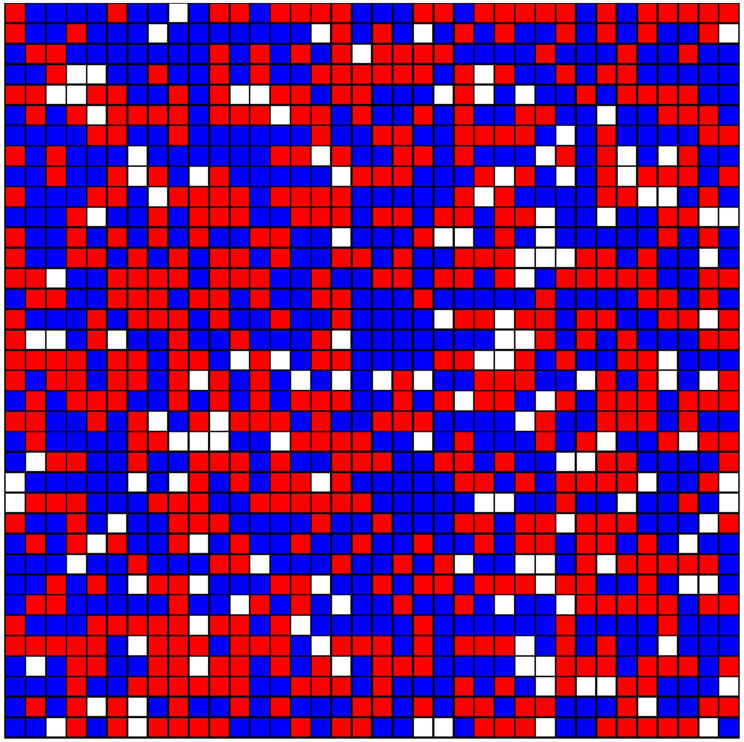
Example: One H_2O molecule -> Wetness of Water

Example for Emergence: Schelling‘s Model

- *The Schelling model consists of a grid of square patches.*
- *Agents are located on this landscape, initially at random, with no more than one on* any patch.

[Schelling, 1969, 1971]

Agents: red or blue squares



Free positions: white squares

Schelling‘s Modell – Live Example

- mild preference of the agent’s own color, results in segregation
  - Each agent has a tolerance parameter. Green agents are “happy” when the ratio of greens to reds in its **Moore neighborhood** (the eight immediately adjacent cells or patches) is more than its tolerance
  - Unhappy agents are allowed to move

http://nifty.stanford.edu/2014/mccown-schelling-model-segregation/

Settings:

Similar 66

Red/Blue: 50/50

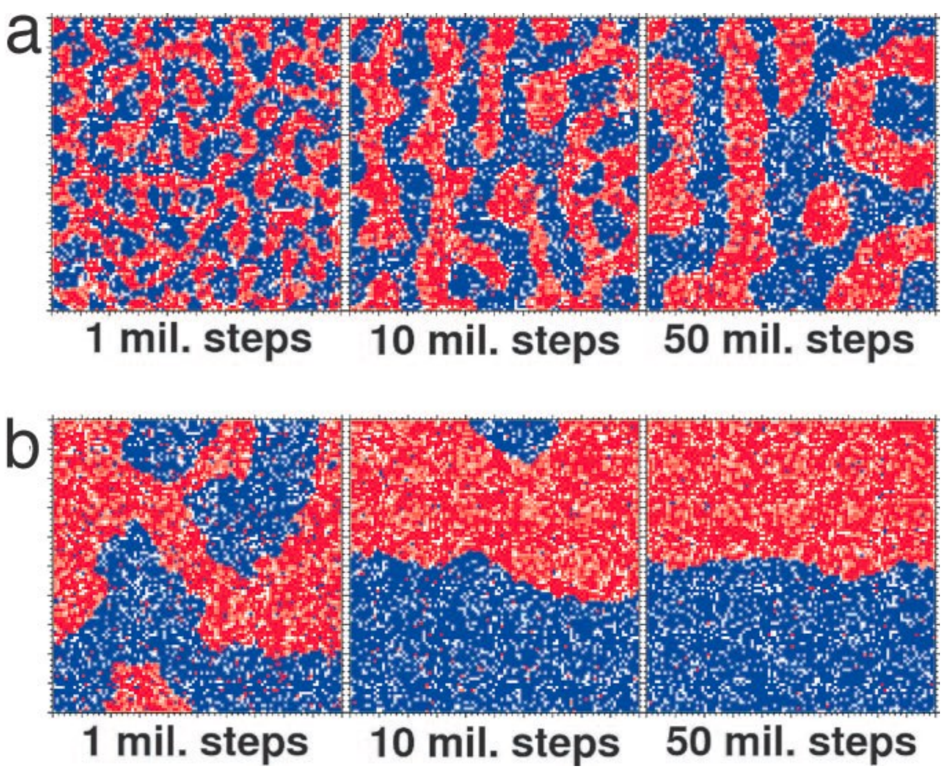
Empty: 10

Size 36x36

Delay 11ms

Schelling's Modell

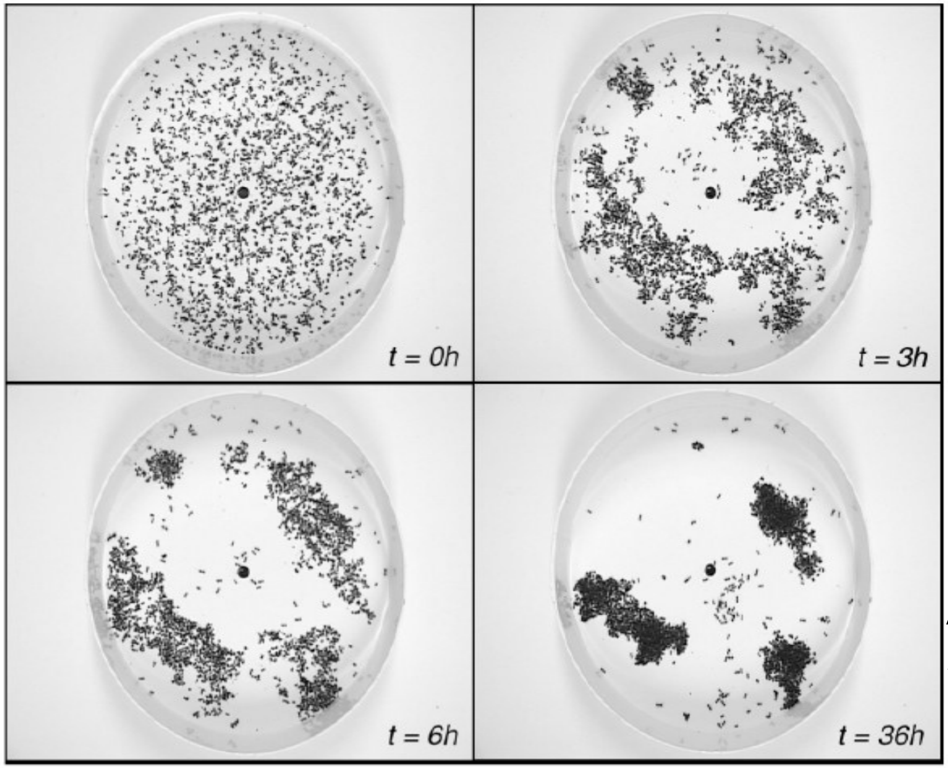
- Behaviour Explanation of emergence of Ghetos:
  - initially integrated communities changed to full segregation
  - even if the people's happiness rules expressed only a mild preference for having neighbors of their own type



Bionics: Ants Swarm Behaviour

- the corpses are randomly distributed at the beginning of the experiment
- Observation: Ant workers pile up their colony's dead to clean their nests.

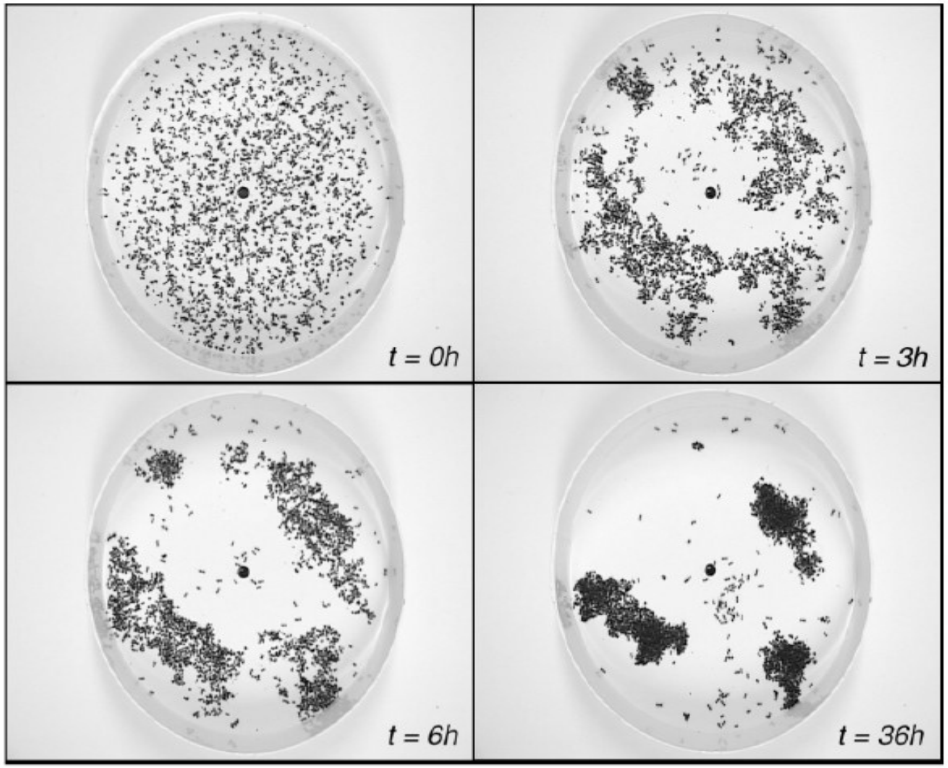
1500 corpses are randomly located in a circular arena.



After 36 hours, well separated clusters have emerged

Example: Ants

- Jean-Louis Deneubourg proposed explanation for the behaviour of the *Messor sancta (1991)*:
  - Small groups of items grow by attracting workers to deposit more items
  - positive feedback leads to the formation of larger and larger bunches



Why do we care about ants?

1. New data can be added incremently
2. Redundant decentralized alogorithm
3. No objective function => possibility of emergence
4. Alternative to neural networks

From: Hermanns WSOM talk

(Mögliche) Schwarmssteme in DataMining

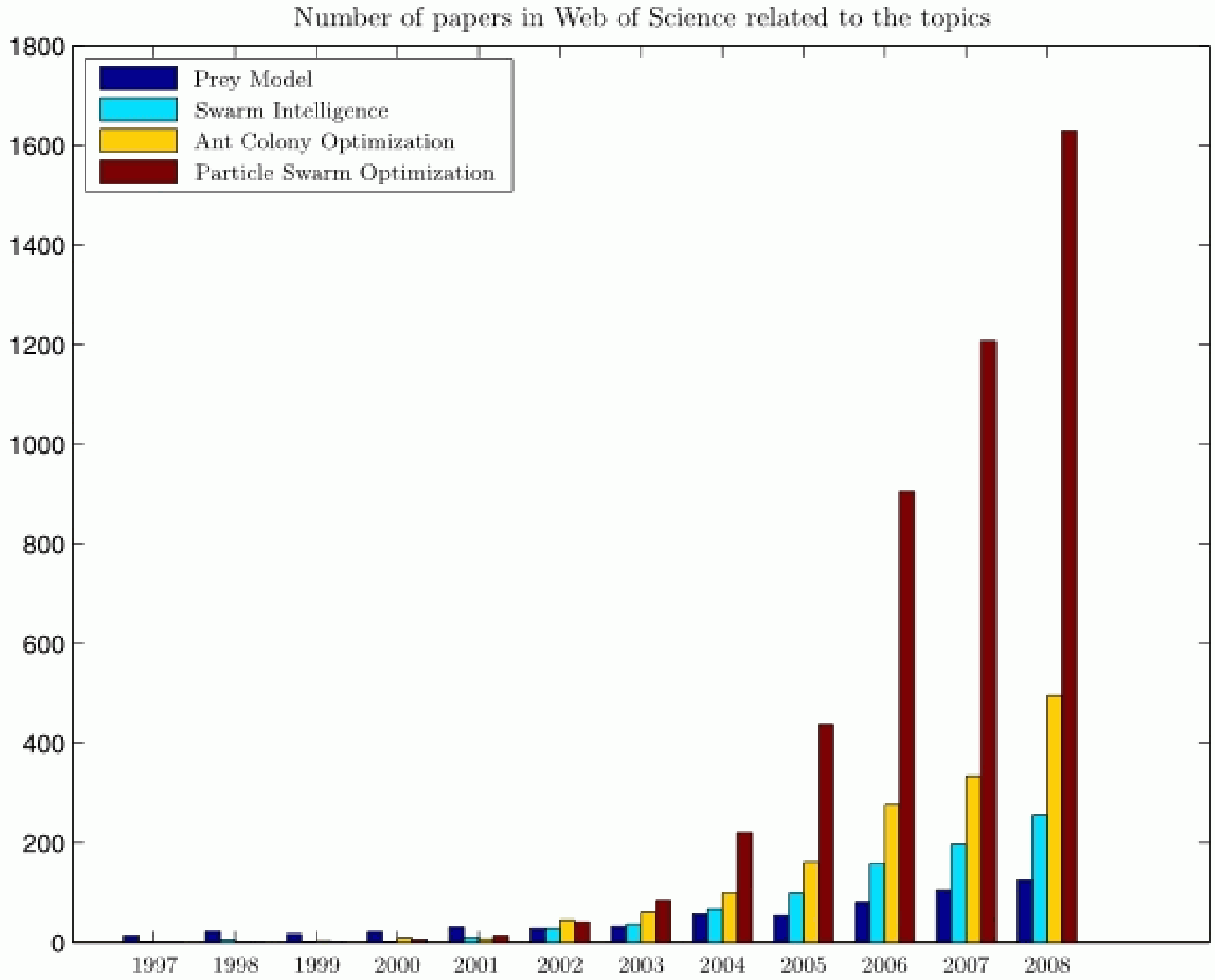


Fig. 1 Number of papers written on the subject, from Web of Science

2012, Martens et al.:

swarm intelligence for data mining

(Mögliche) Schwarmsysteme

Artificial Life (pattern searching):

- Fawcett, Tom. "Data mining with cellular automata." *ACM SIGKDD Explorations Newsletter* 10.1 (2008): 32-39.
- de Buitléir, Amy, Michael Russell, and Mark Daly. "Wains: A pattern-seeking artificial life species." *Artificial life* 18.4 (2012): 399-423.

Ants (Clustering):

- Dorigo, M., et al. (2006): Ant Colony Optimization and Swarm Intelligence. 5th International Workshop, ANTS 2006. Brussels, Belgium, September 4-7, 2006. Proceedings. Springer-Verlag. Berlin Heid

Bird flocking (population-based search algorithm, hybrid for clustering):

- Kennedy, J. and Eberhart, R. C. (2006): Particle swarm optimization. Proc. IEEE int'l conf. on neural networks Vol. IV, pp. 1942-1948. IEEE service center, Piscataway, NJ, 1995. PSO Tutorial found at: http://www.swarmintelligence.org/tutorials.php
- Cui X, Potok TE (2005) Document Clustering Analysis Based on Hybrid PSO+K-means Algorithm, Journal of Computer Sciences (Special Issue), ISSN 1549-3636, pp. 27-33

SI for Unsupervised Machine Learning

1. Particle Swarm Optimization (PSO)
   - Bionics: Bird flocking -> agents communicate directly
   - normally applied as a population-based search algorithm [Rana et al., 2011]
   - rule-based classification models, e.g. AntMiner, or as an optimizer within other learning algorithms
2. Ant Colony Optimization (ACO)
   - agents communicate through stigmergy
   - applied to the task of sorting [Martens et al., 2011]
   - **Ant Based Clustering (ABC)**
3. Artificial Behavior based on DataBots

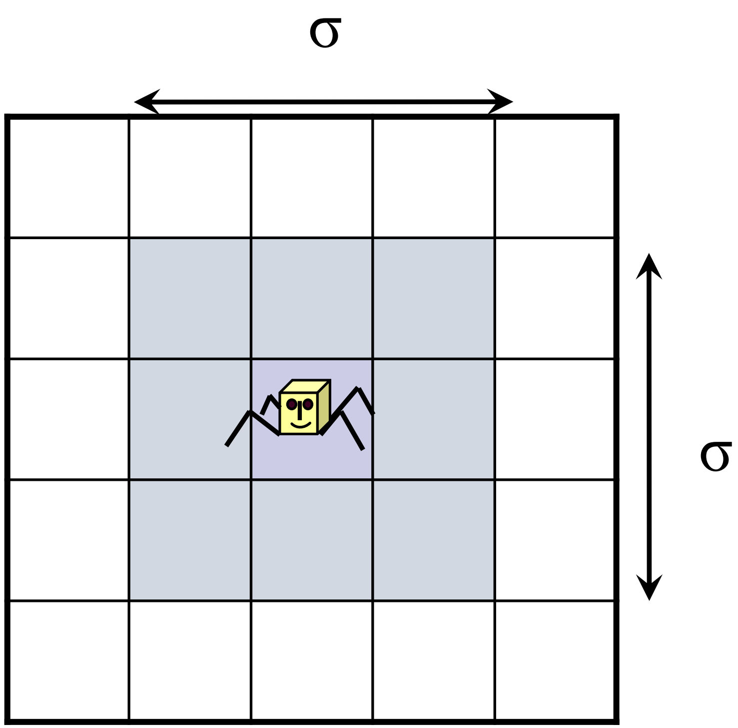
- And some special cases like
  - prey model [Stephens/Krebs, 1986], [Giraldo et al., 2011]

Our Focus: Ant Colonies Optimization (ACO)

- Deneubourg (1990): modelling emergent phenomena of ants clustering corpses
- Lumer, Faieta (1994): clustering pairwise dissimilarity data, e.g. four gaussians
- Ramos (2003): ACLUSTER method includes pheromones, analyzes web traffic data
- Handl, Knowles (2005): ATTA method, first solid empirical evaluations
- Tan et al. (2006): empirical evaluation: number of ants is irrelevant

Ant-Based Clustering (ABC)

Example:
01AntBasedClusteringAnimation.pps



- stochastic ants perform random walks on a grid
- fixed perceptive neighbourhood of size σ^2
- **one or many ants**:
- pick input sample when neighbourhood contains dissimilar samples
- drop input sample when neighbourhood containts similiar samples

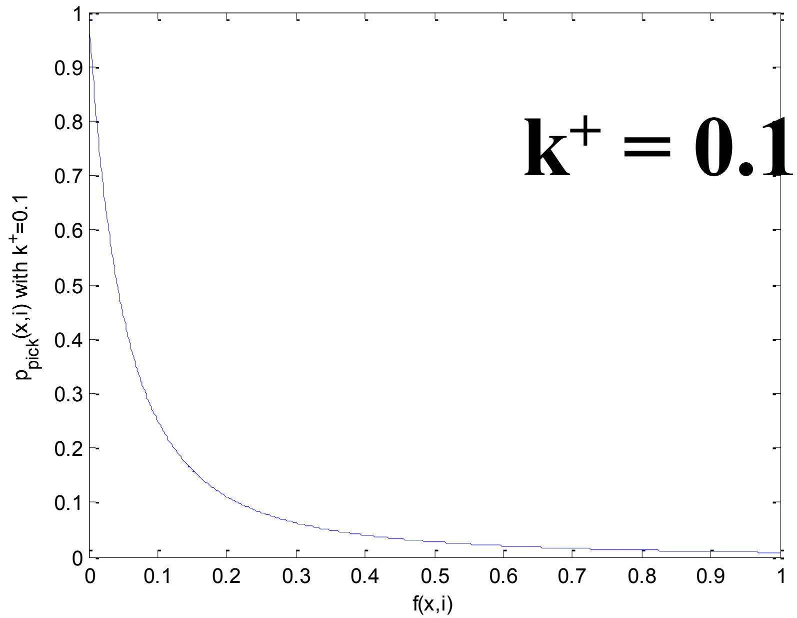
From: Hermanns WSOM talk

Details: Picking

- **The probability *p* for an ant to pick object** $x \in X$ on node i is given by

$$p_{pick}(x,i) = \left(\frac{k^{+}}{k^{+} + f(x,i)} \right)^{2}$$

- where
  - k^{+} is a normalization constant
  - $f(x,i)$ ist die „attractiveness"
- des Ortes i für das Objekt x
- (siehe nächste Folie)

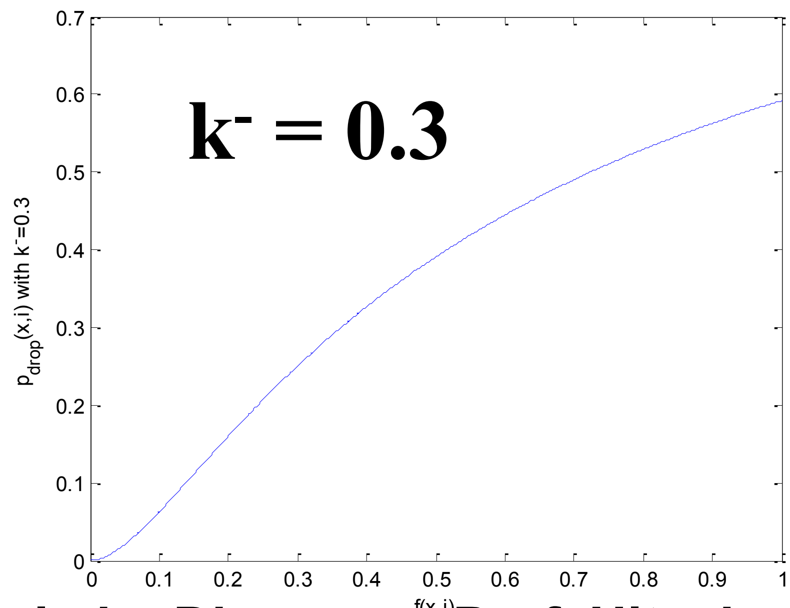


Details: Dropping

- The probability for an ant to drop object $x \in X$ on node i is given by

$$p_{drop}(x,i) = \left(\frac{f(x,i)}{k^- + f(x,i)} \right)^2$$

- where
  - k^- is a normalization constant



From: KnowledgeDiscovery, Prof. Ultsch

Details: Attractiveness function

- The attractiveness for a pick/drop operation of input sample x 2 X on node i is given by

d(.,.) → Ähnlichkeit

Ähnlichkeit pro Fläche

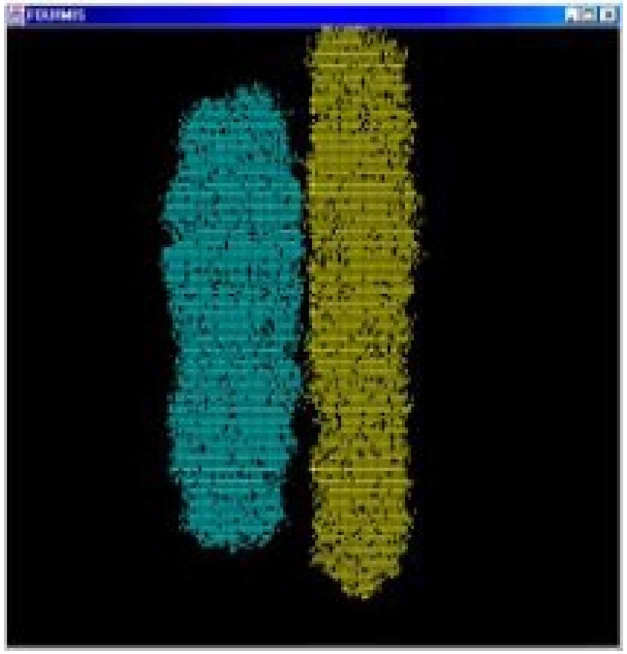
$$f(x,i)=\begin{cases}\frac{1}{\sigma^2}\sum_{y\in N_x(i)}\left(1-\frac{d(x,y)}{\alpha}\right) & >0\\ 0 & else\end{cases}$$

- where
  - σ^2 is the size of the perceptive neighbourhood
  - $N_x(i)$ is the perceptive neighbourhood‘s content
  - α2R is a fixed normalization constant
  - d : X×X /R⁺ is a distance measure

From: KnowledgeDiscovery, Prof. Ultsch

Properties of Ant-Based Clustering

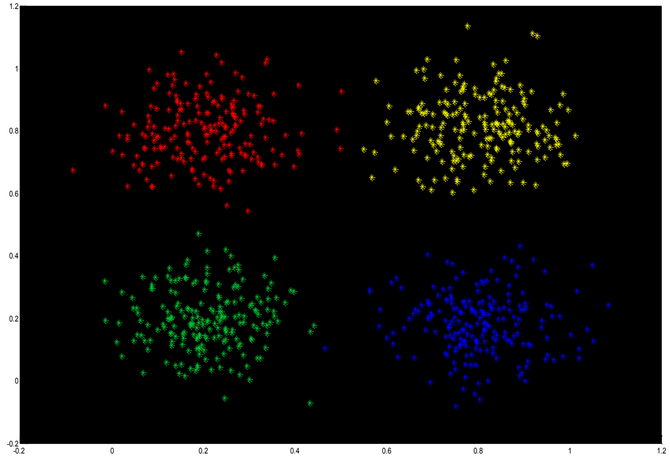
- 20 ants on 200×200 grid
- 5000 yellow & blue objects (one binary attribute)
- 20,000,000 iterations
- [Dorigo 2001]
- Usually, the number of iterations is unacceptably large.



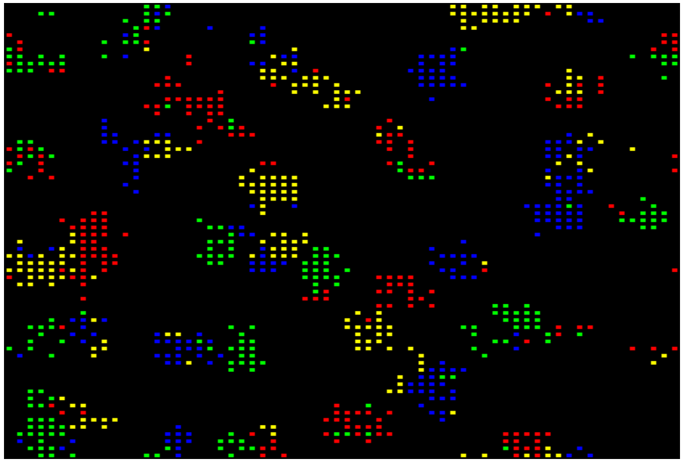
From: KnowledgeDiscovery, Prof. Ultsch

Problems with Ant-Based Clustering

data:
4 gaussian
cluster



data processed with Lumer/Faieta algorithm



- too many, too small clusters appear
- topologically distorted clusters
- results are highly sensitive to parameters, e.g. grid size
- no proof of topological ordering
- Cluster können immer wieder ab und umgebaut werden.

From: Hermanns WSOM talk

Properties of Ant-Based Clustering

- **The original algorithm (LF) does <u>never</u> converge.**
- **Cluster können immer wieder ab und umgebaut werden.**
- **In [Vizine/deCastro 04] a cooling scheme for k^+ was proposed.**
- **The complexity order of LF remains, however, unknown.**

More Observations: Tan et al. (2006):

- Reduce the number of ants to 1
  - Hypothesis: Single ant → no interactions no self-organization
  - Result: Works just as well, no critical mass
  - Shows the principle of **stigmergy** (ant interacts with itself through the environment)

Analysis of ABC by [Hermann 2009]

- **Look at ants as moving DataPoints**

$\Rightarrow$ **Compare ABC with Batch-SOM:**

- Learning scheme of Batch-SOM similar to SOM but all Weights are summed and updated per eppoch instead of updating the weights after each data point:

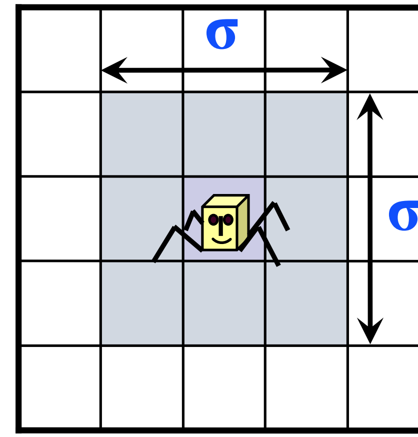
$$w_i \leftarrow \frac{\sum_{y \in X} h_{m(y),i} \cdot y}{\sum_{y \in X} h_{m(y),i}}$$

=> Batch SOM has objective function

$$\Phi_x(i) = \frac{\left\| \sum_{y \in X} h_{m(y),i} \cdot (x - y) \right\|}{\sum_{y \in X} h_{m(y),i}}$$

Details: Attractiveness function for ants

- Define neighborfunction h(σ), which equals the stigmergy of ants:



- Attractiveness function yields to

$$\phi_x(i) = \frac{|N_x(i)|}{\sigma^2}\left(1 - \frac{1}{\alpha} \cdot \frac{\quad}{\sum_{y \in X} h_{m(y),i}}\right)$$

- Where $|N_x(i)|$ is the number of input samples located in the perceptive neighbourhood around node i

Details: Fehlerfunktion für ABC

- Anzahl Punkte in der Nachbarschaft

Groesser als der Topographische Term Φ der Batch-SOM Zielfunktion

$$\phi_x(i) = \frac{|N_x(i)|}{\sigma^2} \cdot \left(1 - \frac{\Phi'_x(i)}{\alpha}\right)$$

- Grösse des sensorischen
- Feldes

- Topographischer Stress wird umgrerechnet in Topograpieerhaltung Alpha = parameter hierfuer

Results of Hermann's Analysis

Ants maximize the product: Output density x topography preservation

- This distorts the formation of a correct data visualization
- objective function includes *optimization of output density*
  - preservation is of doubtful value
  - distorts the topographic term:
    - output space densities easy to optimize and
    - this will dominate the objective function

From Ants to Swarms

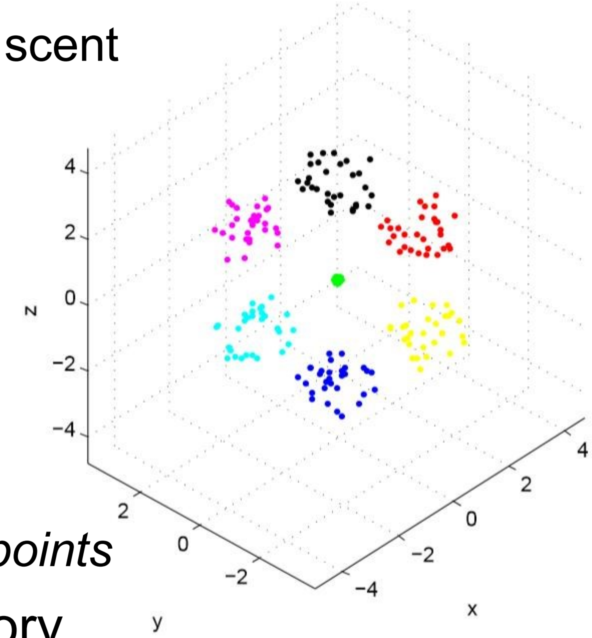
- Ants
  - use small, but fixed neighbourhoods
  - account for output space density
  - One ant sufficient!: no interaction -> no self-organization [Herrmann, 2007,2008,2009]

- Swarms
  - use large, shrinking neighbourhoods [Tsai et al., 2004], [Kämpf/Ultsch, 2006]
  - do not account for output space density [Herrmann, 2009]
  - Many agents [Herrmann, 2009]

Key Idea’s for a swarm

- Swarm intelligence
  1. Required 5 properties
  2. Behavior Modell:
     1. Smelling the surroundings of ones place
     2. “Ant” moves to a free positions, if it prefers the scent of the new position
- Self-organization
  - Required 4 properties
- Application of Bionics (Observation of Ants)
  1. Communication: Scent
  2. Living in and moving on a flat surface
  3. “Ants” wear their colony’s dead -> *will wear data points*
- Use the missing links: Emergence and GameTheory

Goal: separate data into similar groups(“colors”)



How to describe the preference for having neighbors of their own type?

The new approach for a swarm

- Requirements for ***preference***
  - Topoligical neighborhood and distance on grid
  - Correct position determination
- Requirements for projection method
  - Focussing scheme with proper reduction of neighborhood

physics:

- Given problem -> Search first for symmetry
- Solution results in parameter reduction
  - Rmin, Rmax
  - Neighborfunction
  - Grid form and grid distance

Databionic Swarm

- **A projection based clustering method based on self-organizing artificial life forms**
- **The collective is called DataBionic Swarm (DBS)**
  - **Every entity is defined as a DataBot**
- **Every DataBot is represented by exact one data point**
- **DataBots communicate through stigmergy:**
  1. **live on a two-dimensional toroid as an environment**
  2. **are able to smell a position and scent**

Idea from: Herrmann, Ultsch: *An artificial Life Approach for Semi-supervised Learning*, 2007