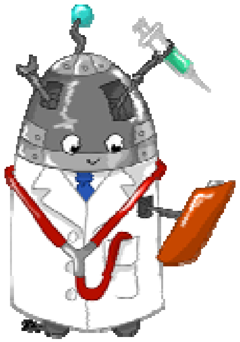
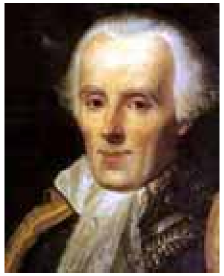


Andreas Holzinger

185.A83 Machine Learning for Health Informatics

2017S, VU, 2.0 h, 3.0 ECTS

Lecture 09 - Module 06 – Week 20 – 16.05.2017



Evolutionary Computing and Agent Interaction (Part 1)

a.holzinger@hci-kdd.org

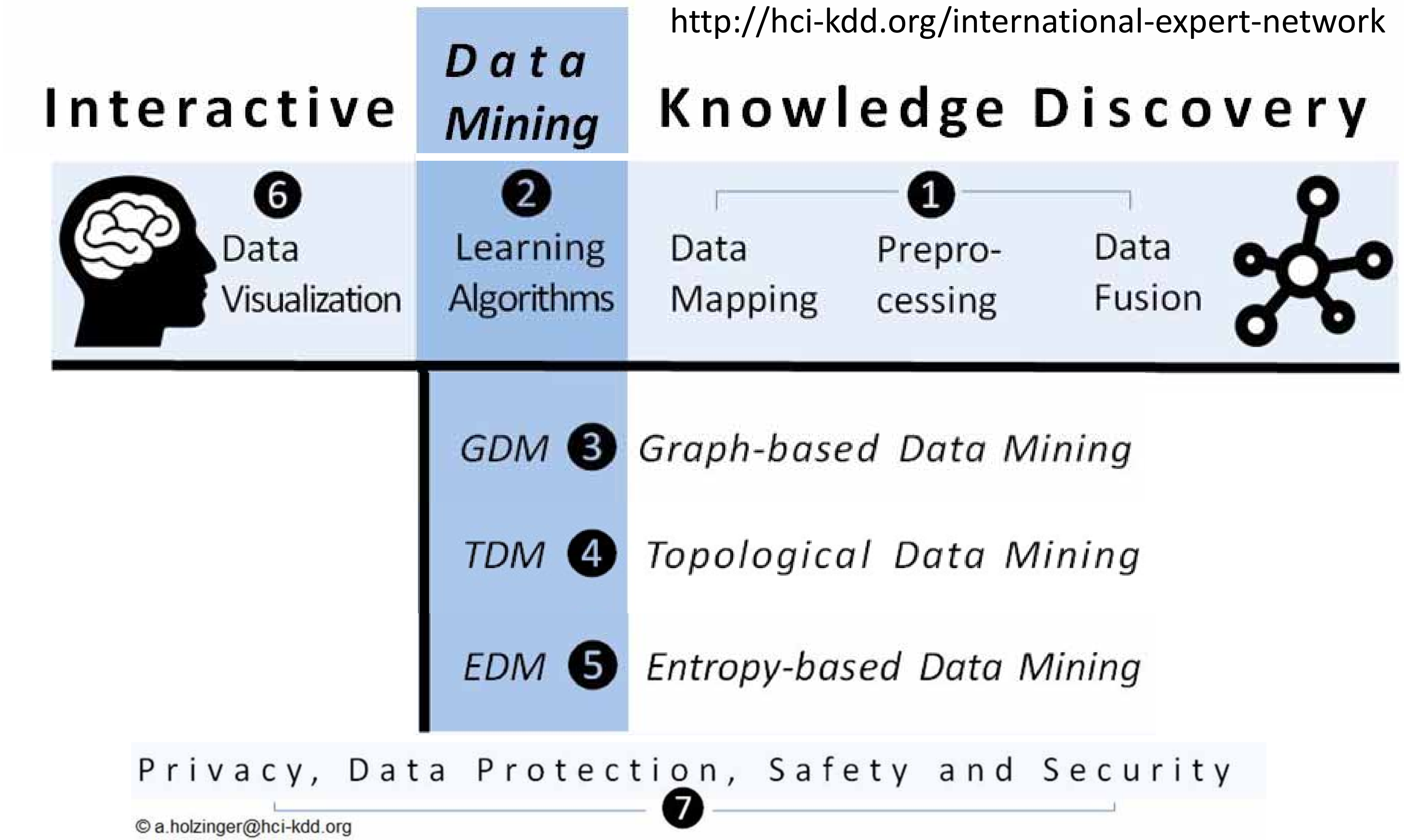
http://hci-kdd.org/machine-learning-for-health-informatics-course

Science **is to test crazy ideas –**

Engineering is to put these ideas into Business

Lucky Students ☺

ML needs a concerted effort fostering integrated research



Holzinger, A. 2014. Trends in Interactive Knowledge Discovery for Personalized Medicine: **Cognitive Science meets Machine Learning.** IEEE Intelligent Informatics Bulletin, 15, (1), 6-14.

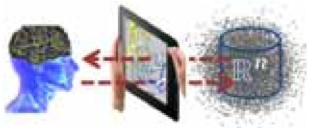
ML-Jungle Top Level View

Maths

Challenges

Cognition | Visualization | Data structure

Perception



Preprocessing

Decision | Interaction | Integration

Always with a focus/application in health informatics

CONCEPTS | THEORIES | PARADIGMS | MODELS | METHODS | TOOLS

Curse of Dim

DR

NfL-Theorem

Overfitting

Non-Parametric

Bayesian p(x)

Complexity

KL-Divergence

Info Theory

unsupervised

supervised

Semi-supv.

online

iML

Gaussian P.

Graphical M.

NN

DL

SVM

Linear Models

D. Trees

RL

PL

AL

Regularization

Validation

Aggregation

Nature Inspired

Python

Julia

Etc.

Azure

Privacy ML

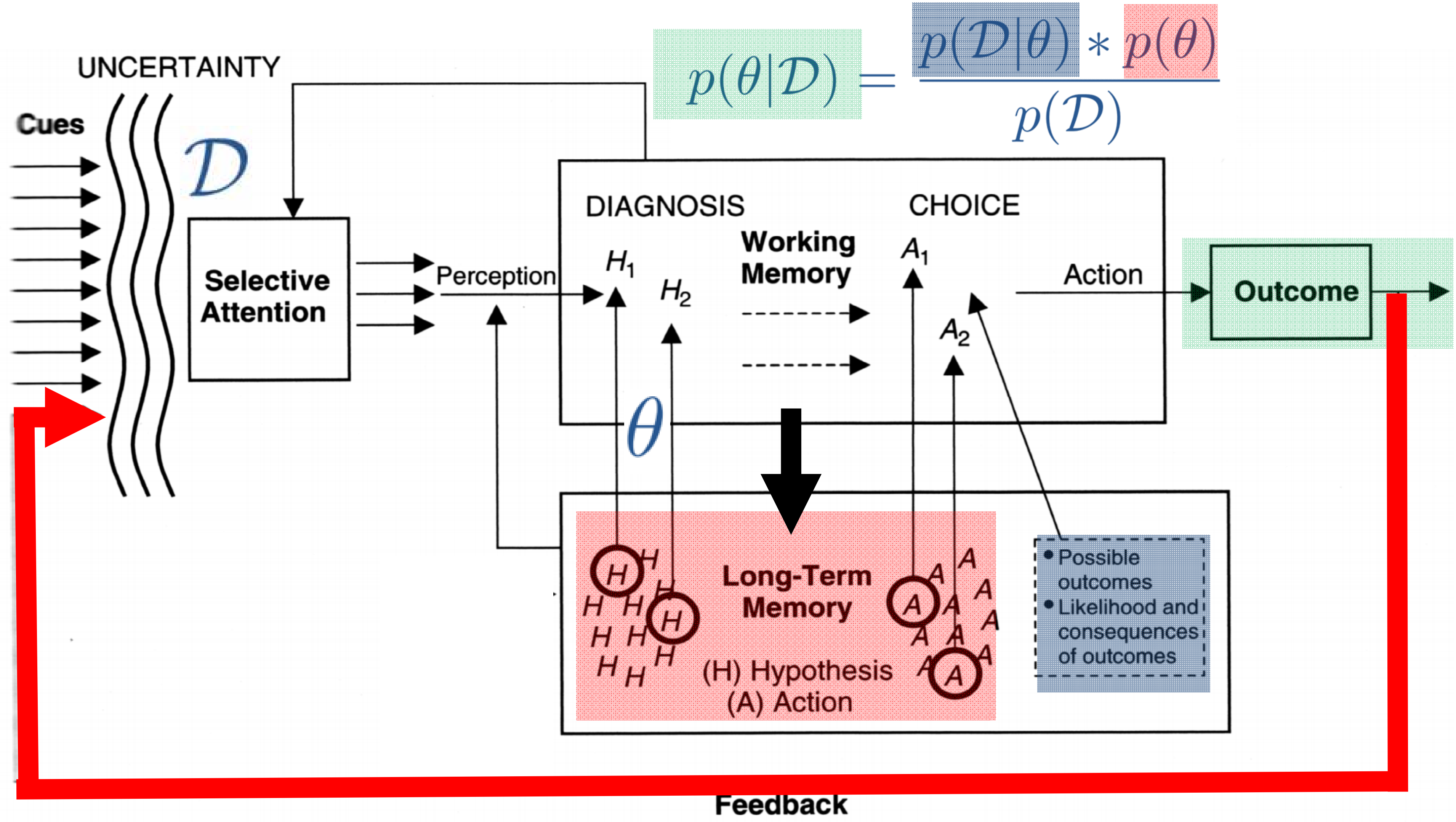
Exp. & Eval.

- 01 Examples of medical applications for EA
- 02 Nature-Inspired Computing
- 03 Ant-Colony Optimization
- 04 Collective Intelligence - Human-in-the-Loop

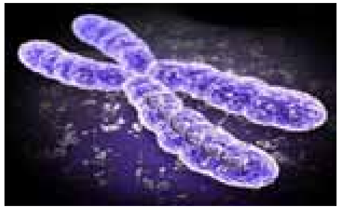
- 05 Multi-Agent (Hybrid) Systems
- 06 Neuroevolution
- 07 Genetic Algorithms

00 Reflection

Human Decision Making: probabilistic reasoning



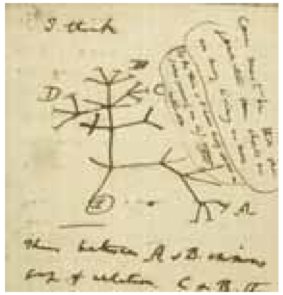
Wickens, C. D. (1984) *Engineering psychology and human performance. Columbus (OH), Charles Merrill, Altered by Holzinger, A. (2017)*



1

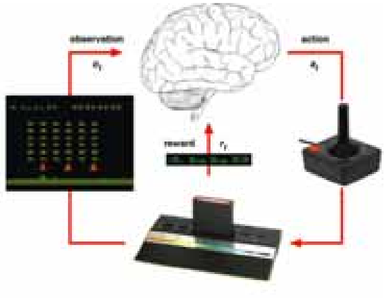
$$V_{wnb}(E) = \arg\max_{c} p(c) \prod_{i=1}^{n} p(a_i|c)^{w_i}$$

2



3

4

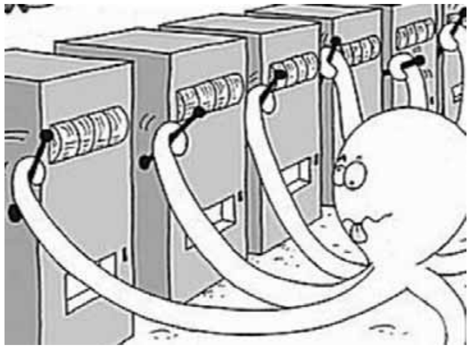


5

10^{80} 10^{40} 10^{120}

4^{3*10^9}

6

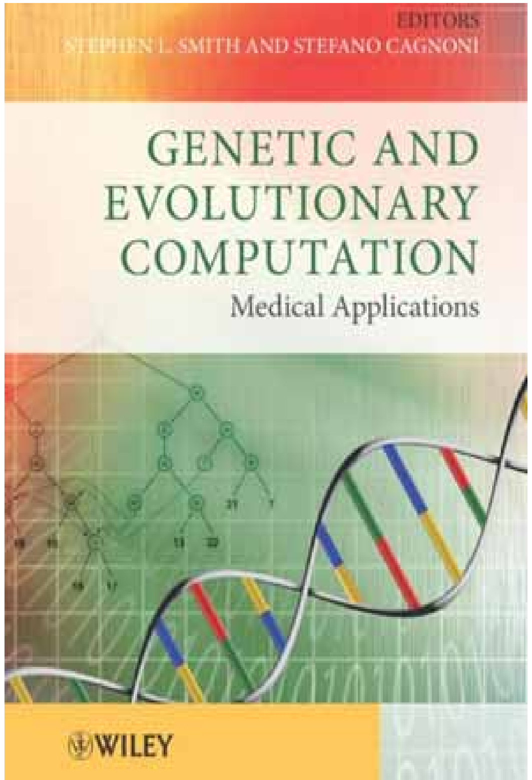


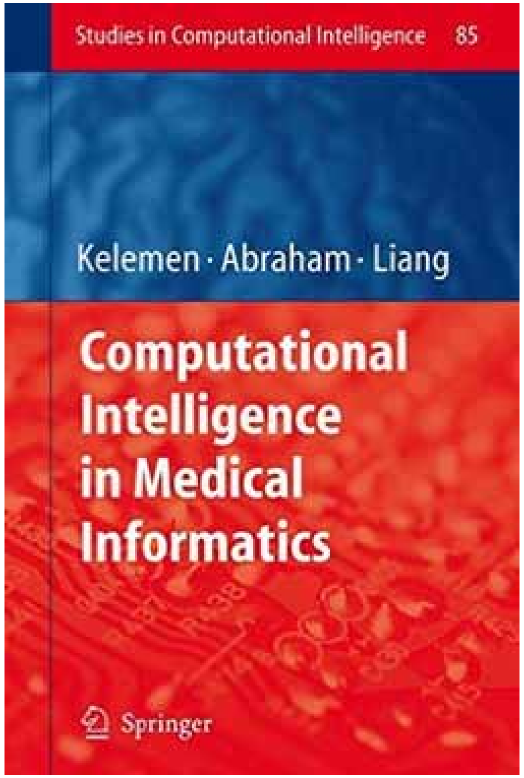
7

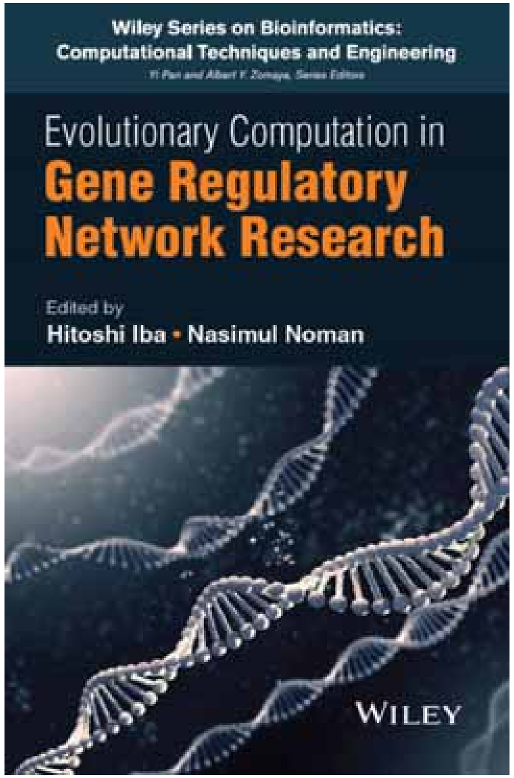
8

- Symbolic ML
  - First order logic, inverse deduction
  - Tom Mitchell, Steve Muggleton, Ross Quinlan, …
- Bayesian ML
  - Statistical learning
  - Judea Pearl, Michael Jordan, David Heckermann, …
- Cognitive ML
  - Analogisms from Psychology, Kernel machines
  - Vladimir Vapnik, Peter Hart, Douglas Hofstaedter, …
- Connectionist ML
  - Neuroscience, Backpropagation
  - Geoffrey Hinton, Yoshua Bengio, Yann LeCun, …
- **Evolutionary ML**
  - Nature-inspired concepts, genetic programming
  - John Holland (1929-2015), John Koza, Hod Lipson, …

01 Applying Evolutionary computation to solve medical problems







Smith, S. L. & Cagnoni, S. 2011. Genetic and evolutionary computation: medical applications, John Wiley & Sons.

Kelemen, A., Abraham, A. & Liang, Y. 2008. Computational intelligence in medical informatics, Springer Science & Business Media.

Iba, H. & Noman, N. 2016. Evolutionary Computation in Gene Regulatory Network Research, John Wiley & Sons.

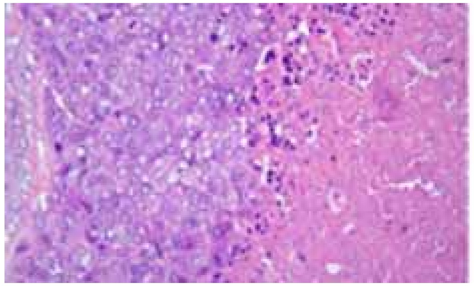
Stephen Smith is at York University (Old York – not New York):
https://scholar.google.at/citations?hl=de&user=T2QamCwAAAAJ&view_op=list_works&sortby=pubdate

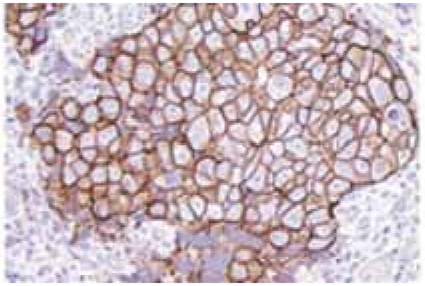
When do we use evolutionary approaches?

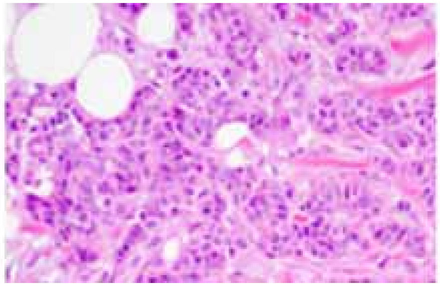
- Many applications in medical imaging, image segmentation, medical data mining, modelling and simulating medical processes, diagnosis, treatment.
- Whenever a **decision** is required, it is possible to find a niche for evolutionary techniques [1]
- Two relevant (and difficult!) questions:
- 1) For a given problem: what is the best algorithm?
- 2) For a given algorithm: what is the problem to solve?

[1] Pena-Reyes, C. A. & Sipper, M. 2000. Evolutionary computation in medicine: an overview. Artificial Intelligence in Medicine, 19, (1), 1-23, doi:10.1016/S0933-3657(99)00047-0.

Example Wisconsin breast cancer diagnosis







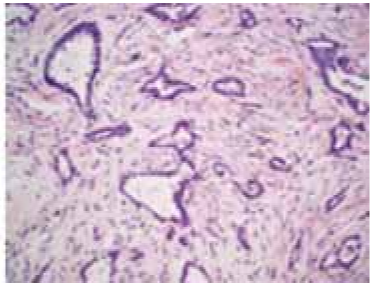


Image Source: https://blogforbreastcancer.wordpress.com/2015/06/30/biopsy-basics-prediction-prognistics-pathology/

```
begin EC
    t:=0
    Initialize population P(t)
    while not done do
        Evaluate P(t)
        P'(t) := Select[P(t)]
        P''(t) := ApplyGeneticOperators[P'(t)]
        P(t+1) := Introduce[P''(t),P(t)]
        t:=t+1
    end while
end EC
```

```
begin GA
    g:=0   { generation counter }
    Initialize population P(g)
    Evaluate population P(g)   { i.e., compute fitness values }
    while not done do
        g:=g+1
        Select P(g) from P(g-1)
        Crossover P(g)
        Mutate P(g)
        Evaluate P(g)
    end while
end GA
```

Pena-Reyes, C. A. & Sipper, M. 2000. Evolutionary computation in medicine: an overview. Artificial Intelligence in Medicine, 19, (1), 1-23, doi:10.1016/S0933-3657(99)00047-0.

Pena-Reyes, C. A. & Sipper, M. 1999. A fuzzy-genetic approach to breast cancer diagnosis. *Artificial intelligence in medicine,* 17, (2), 131-155.

Example: increasing Biploar Disorders (BPD)

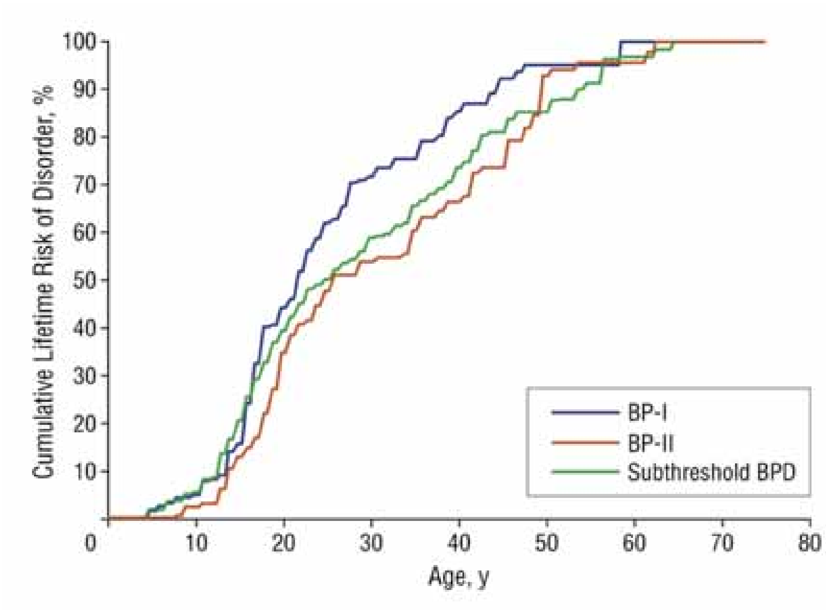


Image credit: http://embracingdepression.org

Table 1. Lifetime and 12-Month Prevalence and Age at Onset of *DSM-IV*/CIDI Bipolar Disorder in the 9282 Respondents

	Any BPD	BP-I	BP-II	Subthreshold BPD
Prevalence, mean (SD)				
Lifetime	4.4 (24.3)	1.0 (13.2)	1.1 (10.6)	2.4 (23.3)
12 mo	2.8 (18.9)	0.6 (9.2)	0.8 (9.9)	1.4 (15.1)
Age at onset, y*				
Mean (SE)	20.8 (11.8)	18.2 (11.6)	20.3 (9.7)	22.2 (12.6)
IQR†	12.6-24.9	12.3-21.2	12.1-24.0	13.0-28.3

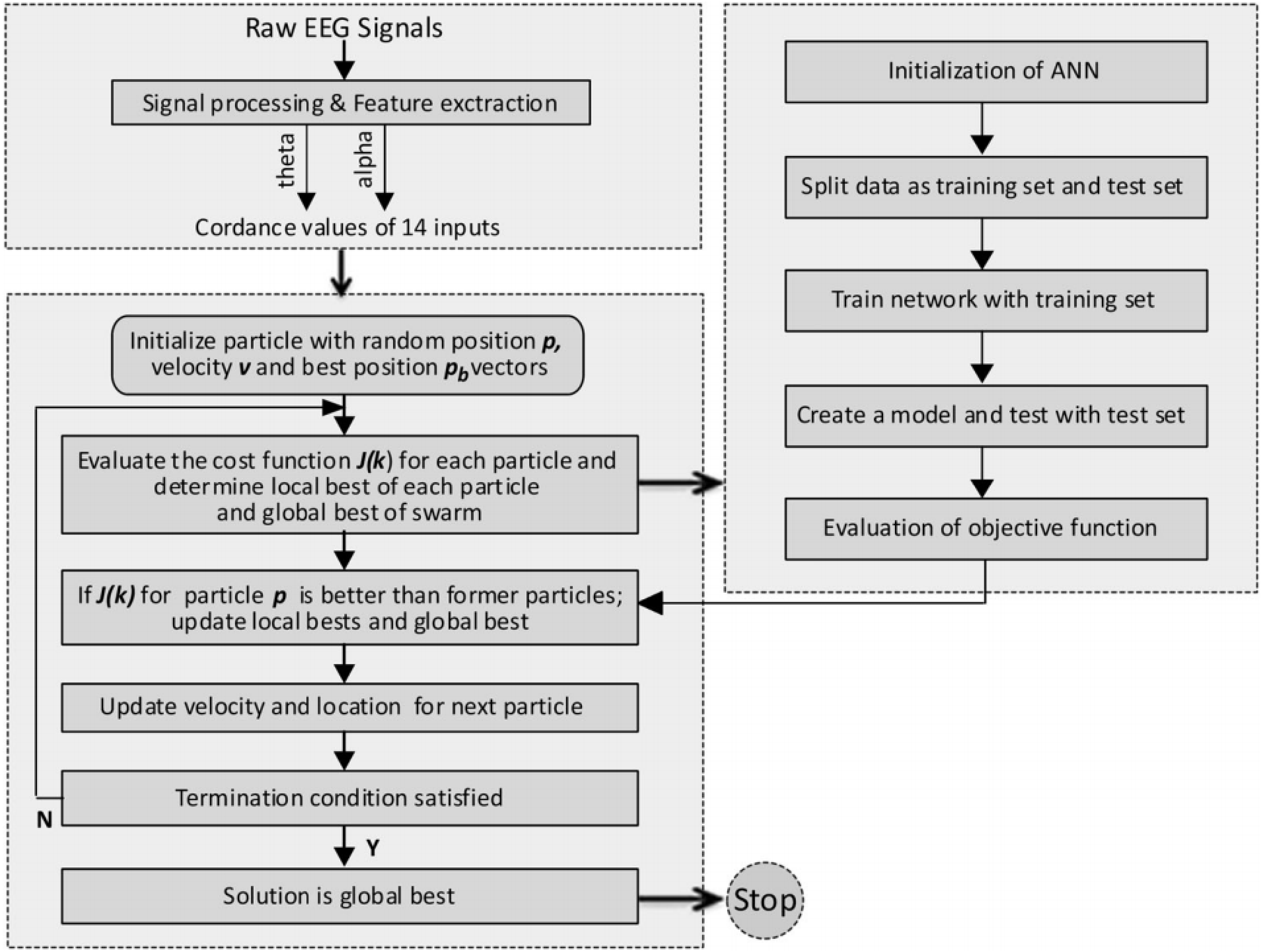
Abbreviations: BPD, bipolar disorder; BP-I, *DSM-IV* bipolar I disorder; BP-II, *DSM-IV* bipolar II disorder; CIDI, Composite International Diagnostic Interview; QR, interquartile range.

*Retrospectively reported age at onset of the first manic/hypomanic or major depressive episode. The means differ significantly across the 3 BPD subgroups at the P=.05 level using a 2-sided test (χ^2_2=7.8; P=.02).

†The range between the 25th and 75th percentiles on the age-at-onset distribution.

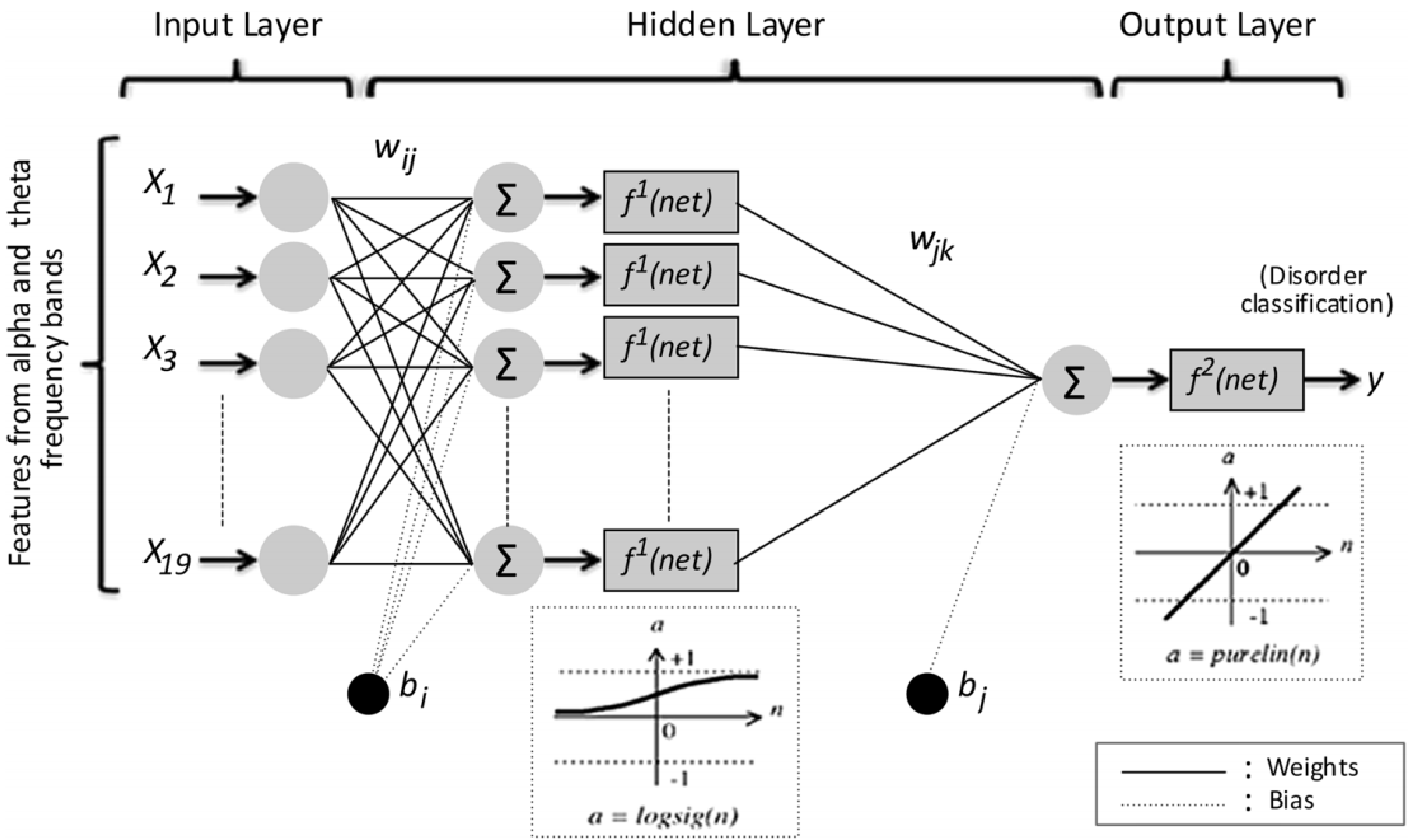
Merikangas, K. R., Akiskal, H. S., Angst, J., Greenberg, P. E., Hirschfeld, R. M., Petukhova, M. & Kessler, R. C. 2007. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Archives of general psychiatry,* 64, (5), 543-552, doi:10.1001/archpsyc.64.5.543.

Feature selection with PSO together with ANN



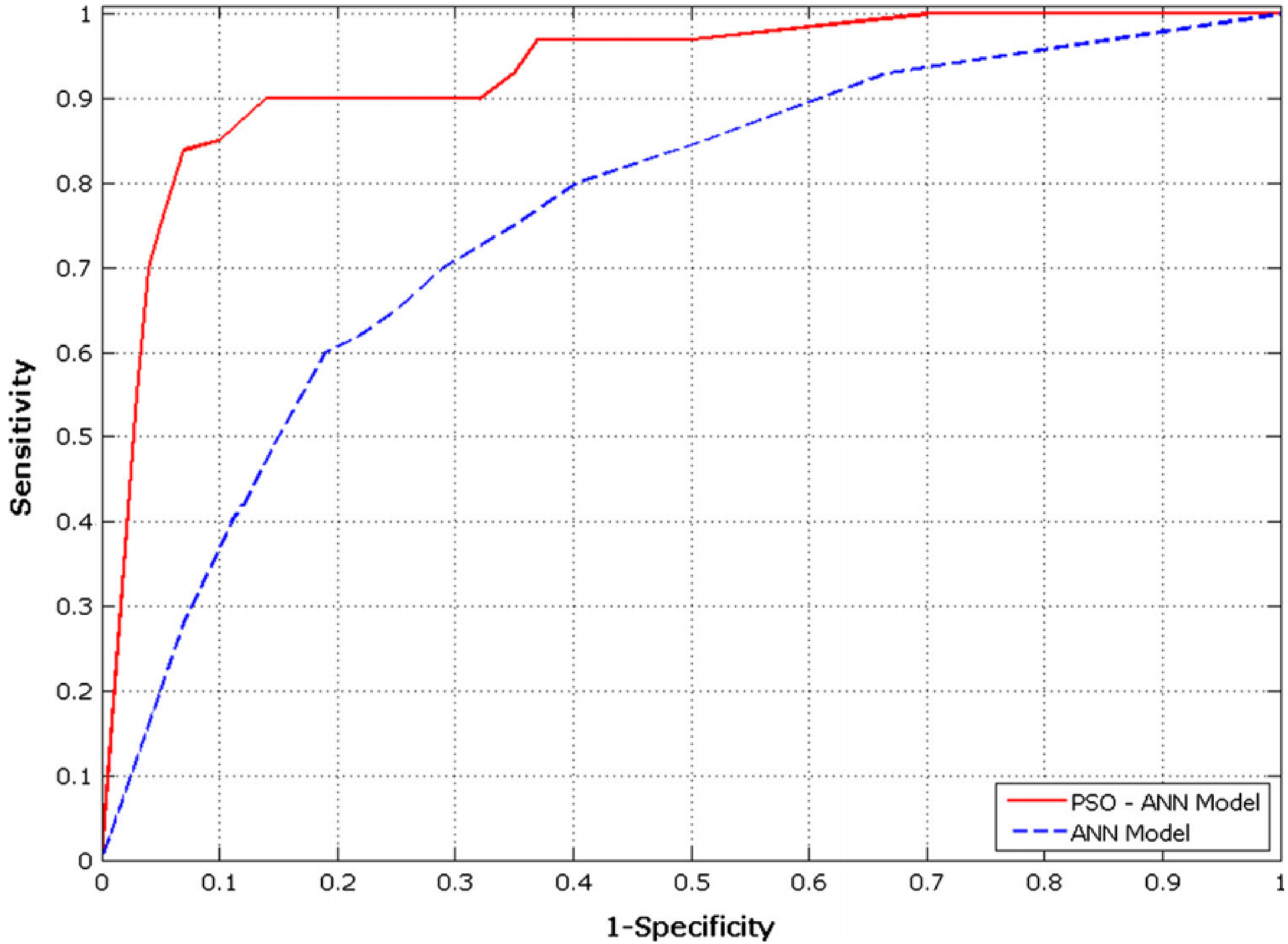
Erguzel, T. T., Sayar, G. H. & Tarhan, N. 2015. Artificial intelligence approach to classify unipolar and bipolar depressive disorders. Neural Computing and Applications, doi:10.1007/s00521-015-1959-z.

Backpropagation ANN

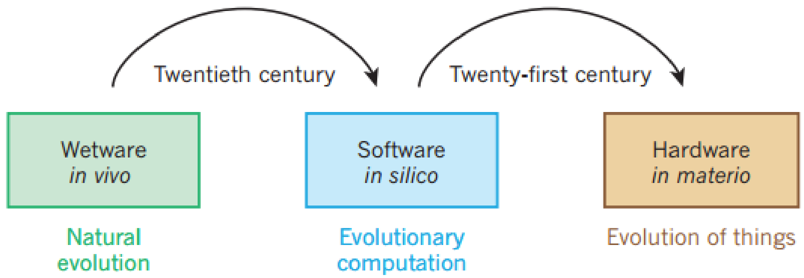


Erguzel, T. T., Sayar, G. H. & Tarhan, N. 2015. Artificial intelligence approach to classify unipolar and bipolar depressive disorders. Neural Computing and Applications, doi:10.1007/s00521-015-1959-z.

Example for PSO: better results than "deep learning"



Erguzel, T. T., Sayar, G. H. & Tarhan, N. 2015. Artificial intelligence approach to classify unipolar and bipolar depressive disorders. Neural Computing and Applications, doi:10.1007/s00521-015-1959-z.

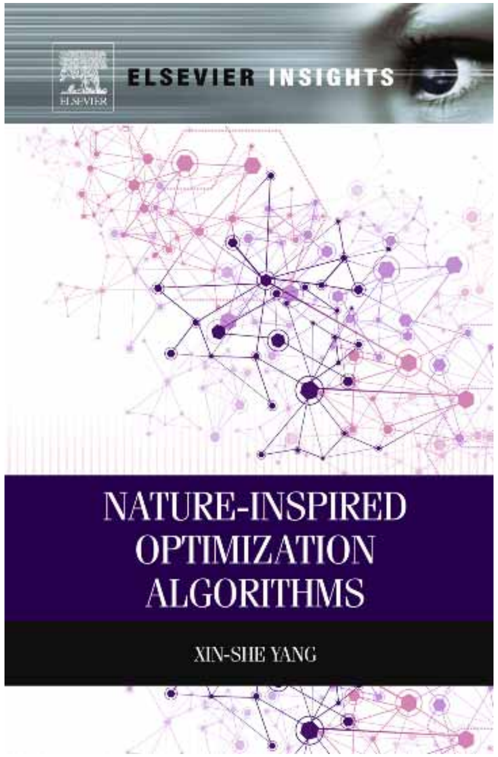


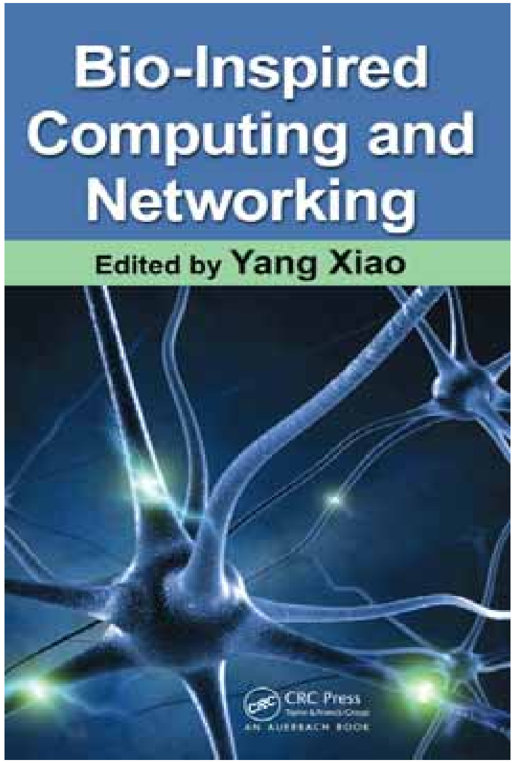
- **Automated design** and tuning of EA for customizing an initial algorithm set-up for a given problem offline (before the run) or online (during the run) and automated parameter tuning
- **Surrogate models:** EA for problems in which evaluating each population member over many generations would take too long to permit effective evolution
- **Multi-objectives** handling at the same time
- **Interactive Evolutionary Algorithms,** bringing in user-preferences, expert knowledge -> human-in-the-loop

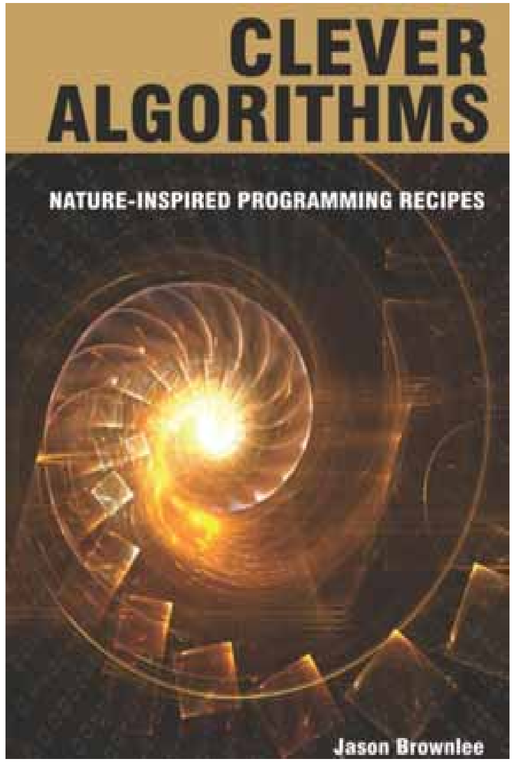
Eiben, A. E. & Smith, J. 2015. From evolutionary computation to the evolution of things. Nature, 521, (7553), 476-482, doi:10.1038/nature14544.

02 Nature Inspired Computing

Recommended Books







Yang, X.-S. 2014. Nature-inspired optimization algorithms, Amsterdam, Elsevier.

Xiao, Y. 2011. Bio-inspired computing and networking, CRC Press.

Brownlee, J. 2011. Clever algorithms: nature-inspired programming recipes, Jason Brownlee.

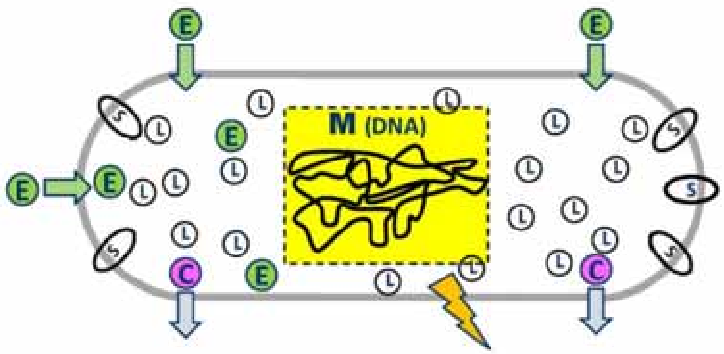
http://machinelearningmastery.com/

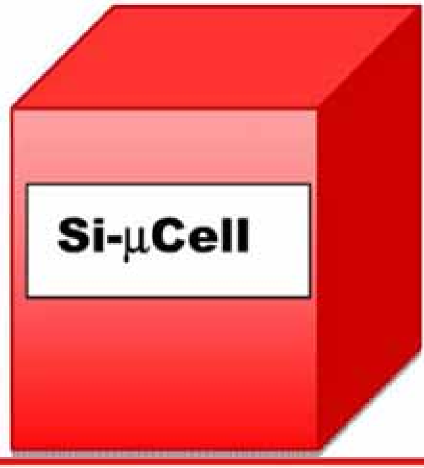
- Computing inspired by phenomena in nature *):
  - Evolutionary Algorithms [1], Genetic Programming etc.
  - Simulated Annealing
  - Swarm Intelligence (Ant, Bee, Bat, Cuckoo, PSO, …)
  - Neuro evolution
  - Random Walks
  - Immuno-computing (Epidemics, Proteins, Viruses, …)
- Simulation/Emulation of Nature
  - Fractals, Cellular automata, Artificial Life
- Natural Computing (with natural materials)
  - Molecular Computing [2]
  - DNA, Membrane (P-Systems) Computing [3]
  - Quantum Computing [4]

[1] Holzinger, K., Palade, V., Rabadan, R. & Holzinger, A. 2014. Darwin or Lamarck? Future Challenges in Evolutionary Algorithms for Knowledge Discovery and Data Mining. In: Lecture Notes in Computer Science LNCS 8401. Heidelberg, Berlin: Springer, pp. 35-56, doi:10.1007/978-3-662-43968-5_3.
[2] Freund, R. & Freund, F. 2001. Molecular computing with generalized homogeneous P-systems. In: Lecture Notes in Computer Science LNCS 2054, Berlin, Heidelberg: Springer, pp. 130-144, doi:10.1007/3-540-44992-2_10.
[3] ppage.psystems.eu (The P-Systems Webpage)
[4] Wittek, P. 2014. Quantum Machine Learning: What Quantum Computing Means to Data Mining, Academic Press.

Why study Natural Computing?

- New forms of synthesizing and understanding nature
- Novel problem solving techniques
- New computing paradigms





Memory:	10^7 bit
Logic:	$>10^6$ bit
Power:	10^{-13} W
Heat:	10^{-6} W/cm^2
Energy/task*:	10^{-10} J
Task time*:	2400s=40min

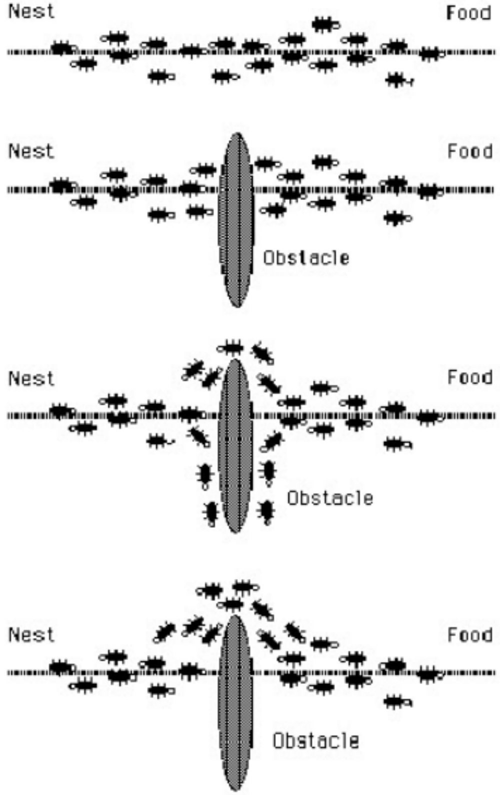
Memory:	~10^4 bit
Logic:	~300–150,000 bit
Power:	~10^{-7} W
Heat:	~1 W/cm^2
Energy/task*:	~10^{-2} J
Task time*:	510,000 s ~ 6 days

*Equivalent to 10^{11} output bits

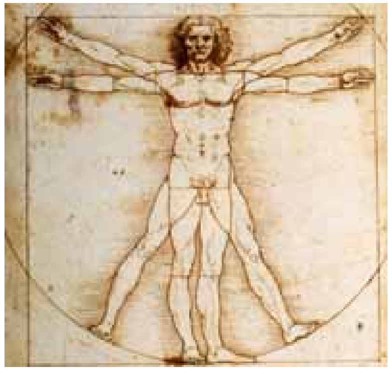
Cavin, R., Lugli, P. & Zhirnov, V. 2012. Science and Engineering Beyond Moore's Law. *Proc. of the IEEE, 100, 1720-49* (L=Logic-Protein; S=Sensor-Protein; C=Signaling-Molecule, E=Glucose-Energy)

Natural Computing Concepts are very useful for us

- Entity (agent)
- Parallelism
- Interactivity
- Connectivity
- Stigmergy *)
- Adaptation
- Feedback
- Self-Organization
- No Self-Organization
- Complexity



*) General mechanism that relates to both individual and colony behaviors – Individual behaviors modify environment – Environment modifies behavior of other individuals – Indirect communication – Example: Ant workers stimulated to act during nest building according to construction of other workers

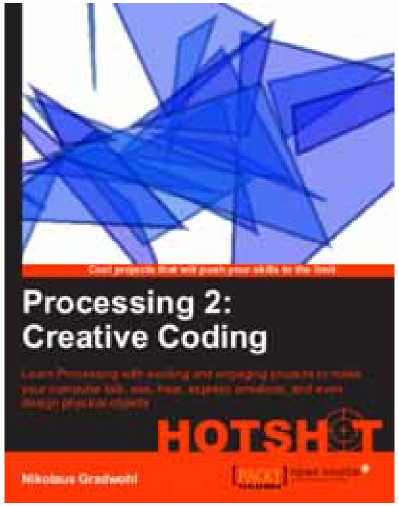


- Population: Collective Intelligence – Swarm Computing (Crowdsourcing HiL)
- Population: Individual – Artificial Life
- Population: Intra-Individual – Evolutionary Computing
- Individual: Neural Networks (Deep Learning)
- Individual: Intra-Individual – Immuno-Computing
- Molecules: Molecular Computing, Biocomputing
- Atoms: Simulated Annealing
- Subatomic: Quantum Computing

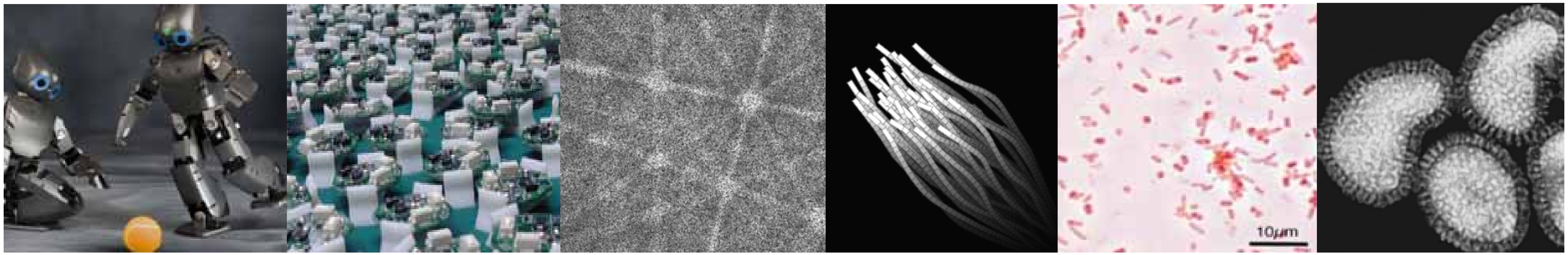
Natural Computing Concept: Entity

https://www.packtpub.com/application-development/processing-2-creative-coding-hotshot

Nikolaus Gradwohl: http://www.local-guru.net/



- Entity (we call it agent later ☺)
  - Acting autonomously, communicating
  - e.g. robots, agents, noise patterns, boids, bacteria, viruses, …, any physical, biological, chemical entity, …



Example: Game of Life, John H. Conway (1970)

https://www.youtube.com/watch?v=CgOcEZinQ2I

http://ddi.cs.uni-potsdam.de/HyFISCH/Produzieren/lis_projekt/proj_gamelife/ConwayScientificAmerican.htm

https://www.youtube.com/watch?v=xbTQ4tqVdz8

http://www.bitstorm.org/gameoflife/

A brief overview of (some) nature inspired algorithms …

Swarm intelligence based algorithms			Bio-inspired (not SI-based) algorithms		
Algorithm	Author	Reference	Algorithm	Author	Reference
Accelerated PSO	Yang et al.	[69], [71]	Atmosphere clouds model	Yan and Hao	[67]
Ant colony optimization	Dorigo	[15]	Biogeography-based optimization	Simon	[56]
Artificial bee colony	Karaboga and Basturk	[31]	Brain Storm Optimization	Shi	[55]
Bacterial foraging	Passino	[46]	Differential evolution	Storn and Price	[57]
Bacterial-GA Foraging	Chen et al.	[6]	Dolphin echolocation	Kaveh and Farhoudi	[33]
Bat algorithm	Yang	[78]	Japanese tree frogs calling	Hernández and Blum	[28]
Bee colony optimization	Teodorović and Dell'Orco	[62]	Eco-inspired evolutionary algorithm	Parpinelli and Lopes	[45]
Bee system	Lucic and Teodorovic	[40]	Egyptian Vulture	Sur et al.	[59]
BeeHive	Wedde et al.	[65]	Fish-school Search	Lima et al.	[14], [3]
Wolf search	Tang et al.	[61]	Flower pollination algorithm	Yang	[72], [76]
Bees algorithms	Pham et al.	[47]	Gene expression	Ferreira	[19]
Bees swarm optimization	Drias et al.	[16]	Great salmon run	Mozaffari	[43]
Bumblebees	Comellas and Martinez	[12]	Group search optimizer	He et al.	[26]
Cat swarm	Chu et al.	[7]	Human-Inspired Algorithm	Zhang et al.	[80]
Consultant-guided search	Iordache	[29]	Invasive weed optimization	Mehrabian and Lucas	[42]
Cuckoo search	Yang and Deb	[74]	Marriage in honey bees	Abbass	[1]
Eagle strategy	Yang and Deb	[75]	OptBees	Maia et al.	[41]
Fast bacterial swarming algorithm	Chu et al.	[8]	Paddy Field Algorithm	Premaratne et al.	[48]
Firefly algorithm	Yang	[70]	Roach infestation algorithm	Havens	[25]
Fish swarm/school	Li et al.	[39]	Queen-bee evolution	Jung	[30]
Good lattice swarm optimization	Su et al.	[58]	Shuffled frog leaping algorithm	Eusuff and Lansey	[18]
Glowworm swarm optimization	Krishnanand and Ghose	[37], [38]	Termite colony optimization	Hedayatzadeh et al.	[27]
Hierarchical swarm model	Chen et al.	[5]	Physics and Chemistry based algorithms		
Krill Herd	Gandomi and Alavi	[22]	Big bang-big Crunch	Zandi et al.	[79]
Monkey search	Mucherino and Seref	[44]	Black hole	Hatamlou	[24]
Particle swarm algorithm	Kennedy and Eberhart	[35]	Central force optimization	Formato	[21]
Virtual ant algorithm	Yang	[77]	Charged system search	Kaveh and Talatahari	[34]
Virtual bees	Yang	[68]	Electro-magnetism optimization	Cuevas et al.	[13]
Weightless Swarm Algorithm	Ting et al.	[63]	Galaxy-based search algorithm	Shah-Hosseini	[53]
Other algorithms			Gravitational search	Rashedi et al.	[50]
Anarchic society optimization	Shayeghi and Dadashpour	[54]	Harmony search	Geem et al.	[23]
Artificial cooperative search	Civicioglu	[9]	Intelligent water drop	Shah-Hosseini	[52]
Backtracking optimization search	Civicioglu	[11]	River formation dynamics	Rabanal et al.	[49]
Differential search algorithm	Civicioglu	[10]	Self-propelled particles	Vicsek	[64]
Grammatical evolution	Ryan et al.	[51]	Simulated annealing	Kirkpatrick et al.	[36]
Imperialist competitive algorithm	Atashpaz-Gargari and Lucas	[2]	Stochastic difusion search	Bishop	[4]
League championship algorithm	Kashan	[32]	Spiral optimization	Tamura and Yasuda	[60]
Social emotional optimization	Xu et al.	[66]	Water cycle algorithm	Eskandar et al.	[17]

Fister Jr, I., Yang, X.-S., Fister, I., Brest, J. & Fister, D. 2013. A brief review of nature-inspired algorithms for optimization. arXiv preprint arXiv:1307.4186.

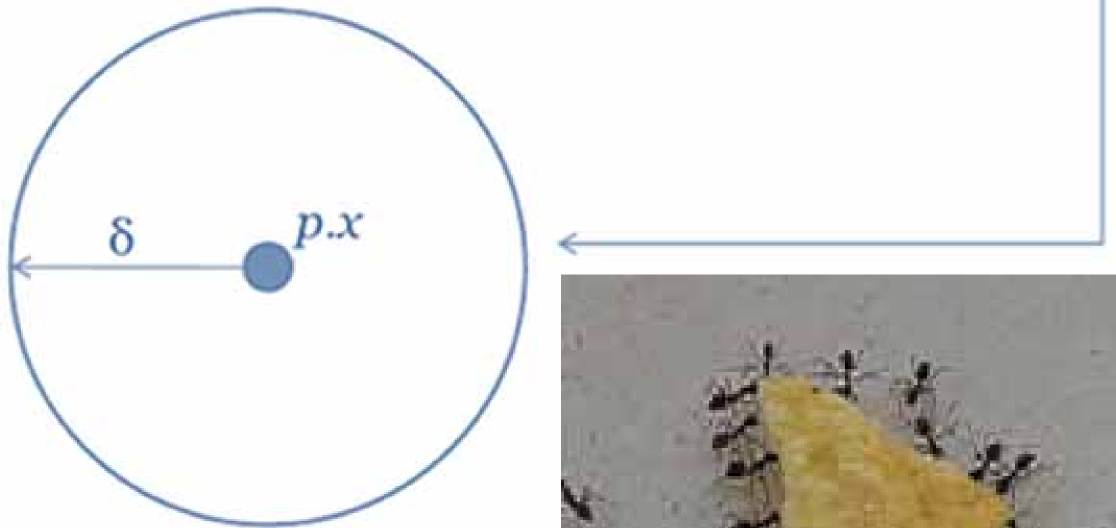
- Particle Swarm Optimization (PSO)
  - based on social behaviour of bird flocks used as method for continuous optimization problems
- Artificial Bee Colonies (ABC)
  - Algorithms based on foraging of honey bee swarms used for continuous optimization problems
- **Ant Colony Optimization (ACO)**
  - Algorithms based on social behaviour of ants, used as metaheuristic for (hard) combinatorial optimization problems (e.g. for TSP-like problems)

How to find good solutions?

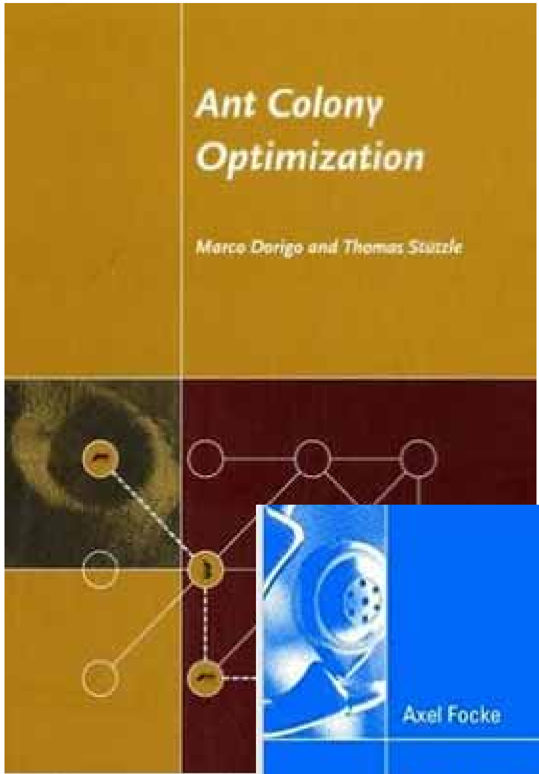
$$\forall p,q \in \boxed{Pop} : q \in \boxed{N(p)} \Leftrightarrow dist_{eucl}(p.g, q.g) \leq \boxed{\delta}$$

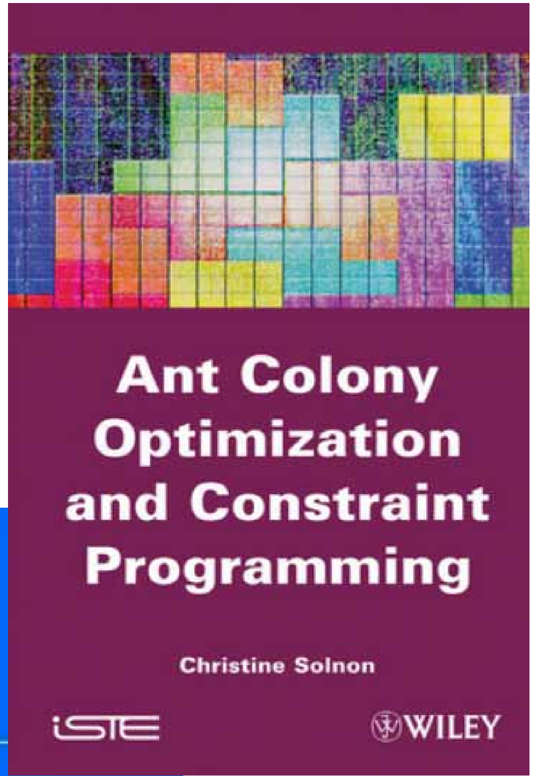
Population

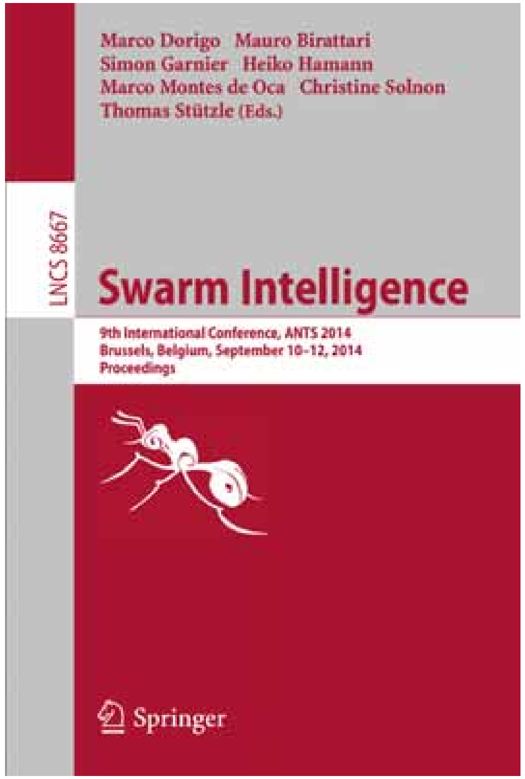
Topological Neighbours



03 Ant Colony Algortihms ACO







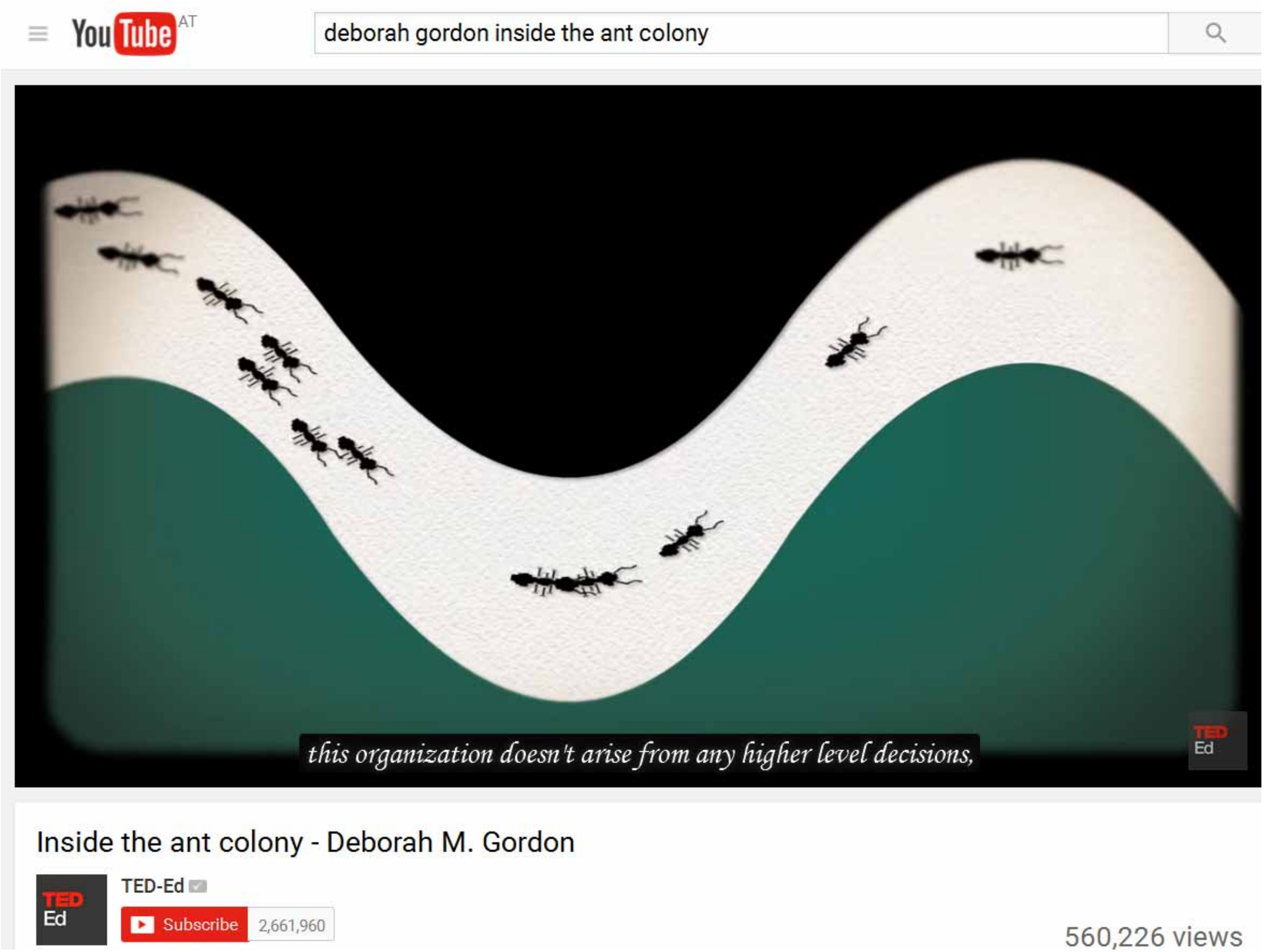
http://alife.org/conference/ants-2016

Dorigo, M., Birattari, M., Blum, C., Clerc, M., Stützle, T. & Winfield, A. 2008. Ant Colony Optimization and Swarm Intelligence: 6th International Conference, ANTS 2008, Brussels, Belgium, September 22-24, 2008, Proceedings, Springer.

Ants as a inspiration for collective intelligence

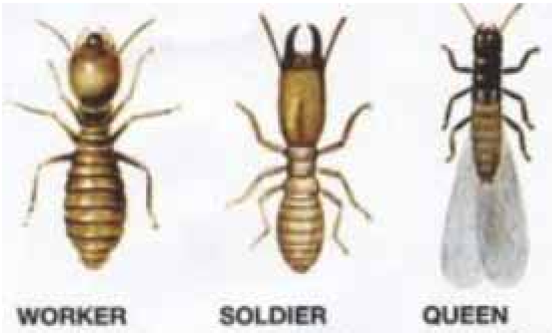
http://www.kurzweilai.net/army-ants-living-bridges-suggest-collective-intelligence

Ant colonies are extremely interesting …



http://web.stanford.edu/~dmgordon/

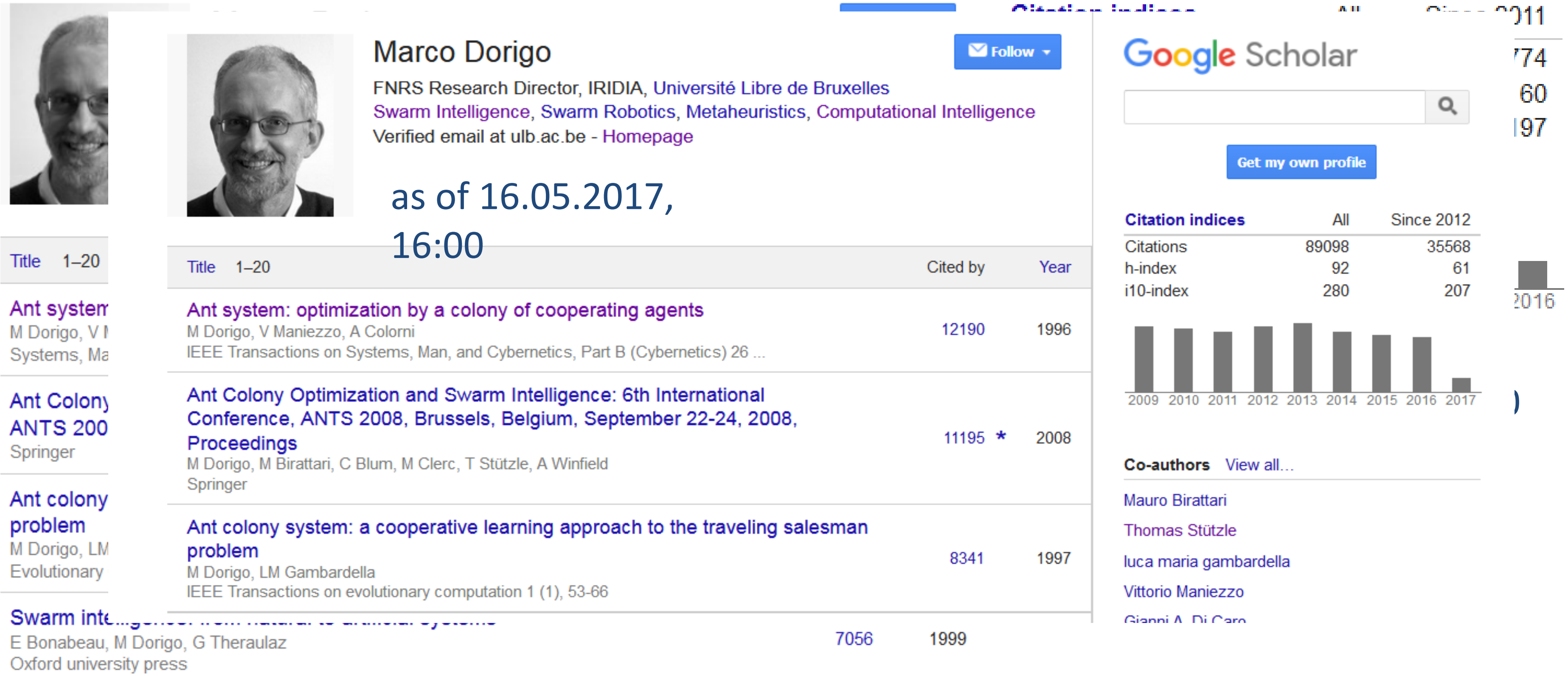
Examples of social intelligent insects:
- Ants
- Termites
- Bees
- Wasps, etc



Some facts:
- 2% of all insects are social
- 50% of all social insects are ants
- Total weight of ants is about the total weight of humans
- Ants colonize world since 100 M years !!! humans only 5 M years …

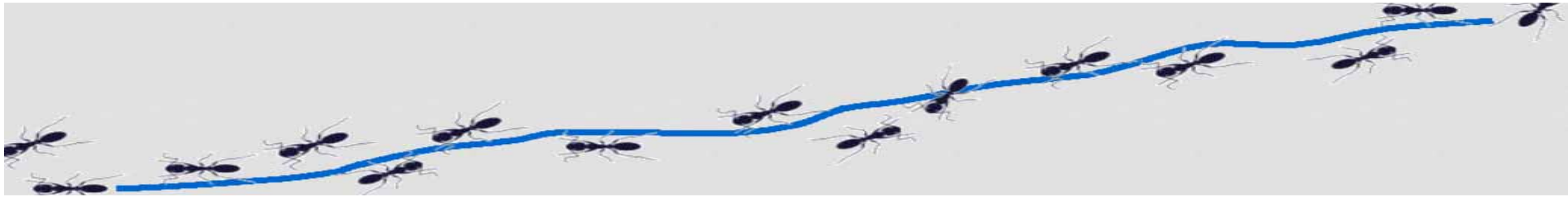
Thanks to the LIACS Natural Computing Group Leiden University

Ant Colony Optimization (ACO) by Marco Dorigo



- Probabilistic optimization inspired by interaction of ants in nature.
- Individual ants are blind and dumb, but ant colonies show complex and smart behavior as a result of low-level based communications.
- Useful for computational problems which can be reduced to **finding good paths in graphs.**

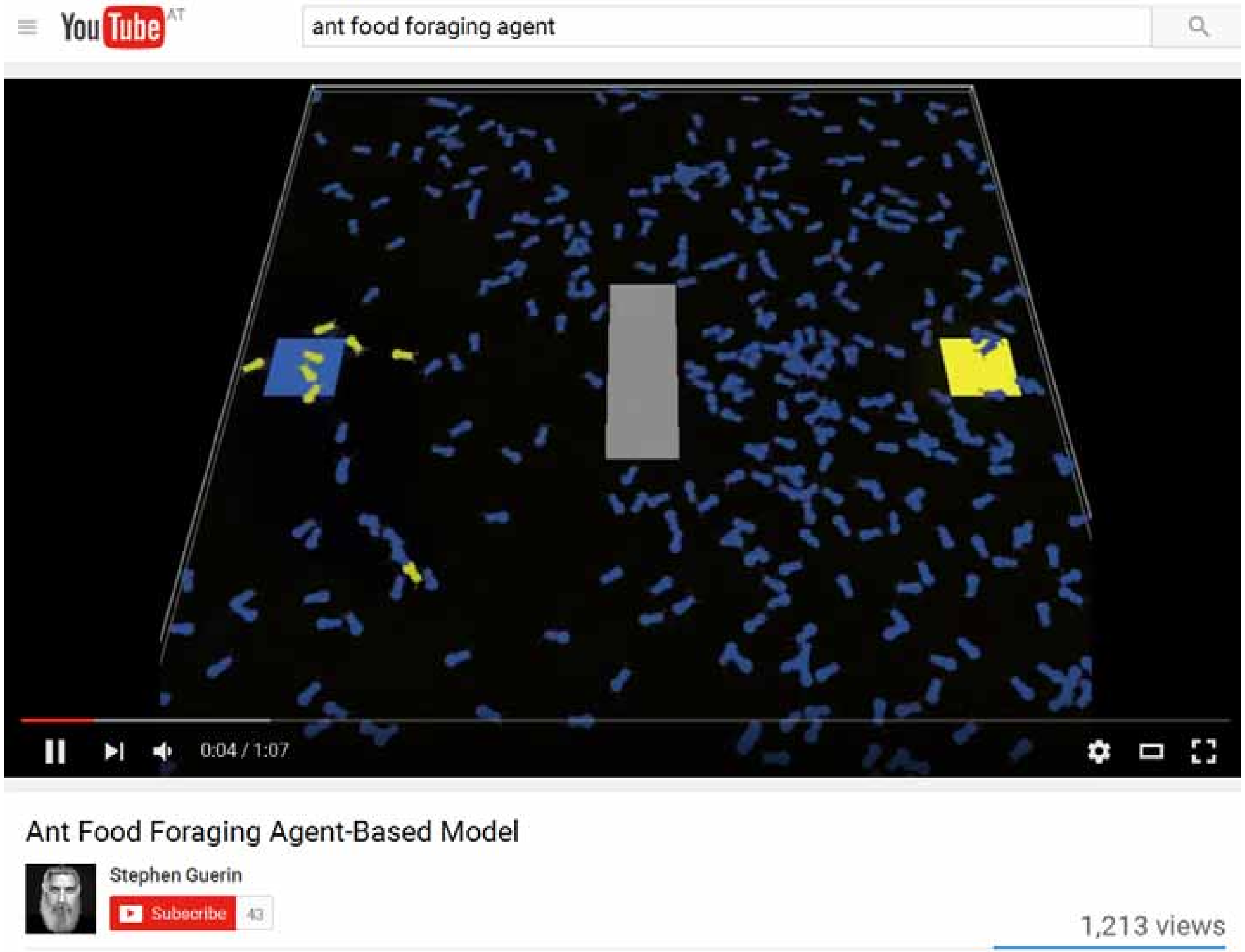
http://iridia.ulb.ac.be/~mdorigo/HomePageDorigo/

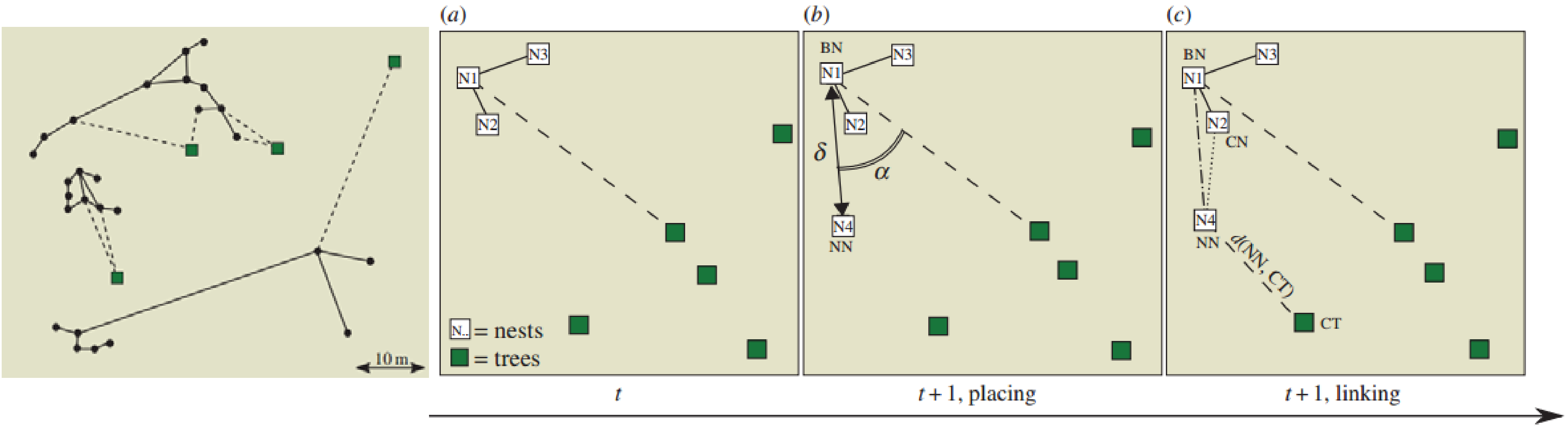


- Ants wander randomly and search for food
- If an ant finds food it returns home laying down a **pheromone trail** on its way back
- Other ants stumble upon the trail and start following this pheromone trail
- Other ants also return home and also deposit pheromones on their way back (**reinforcing** the trail) – when a path is blocked they explore alternative routes …

Colorni, A., Dorigo, M. & Maniezzo, V. 1991. Distributed optimization by ant colonies. Proceedings of the first European conference on artificial life ECAL 91, 134-142.

Ant food foraging agent





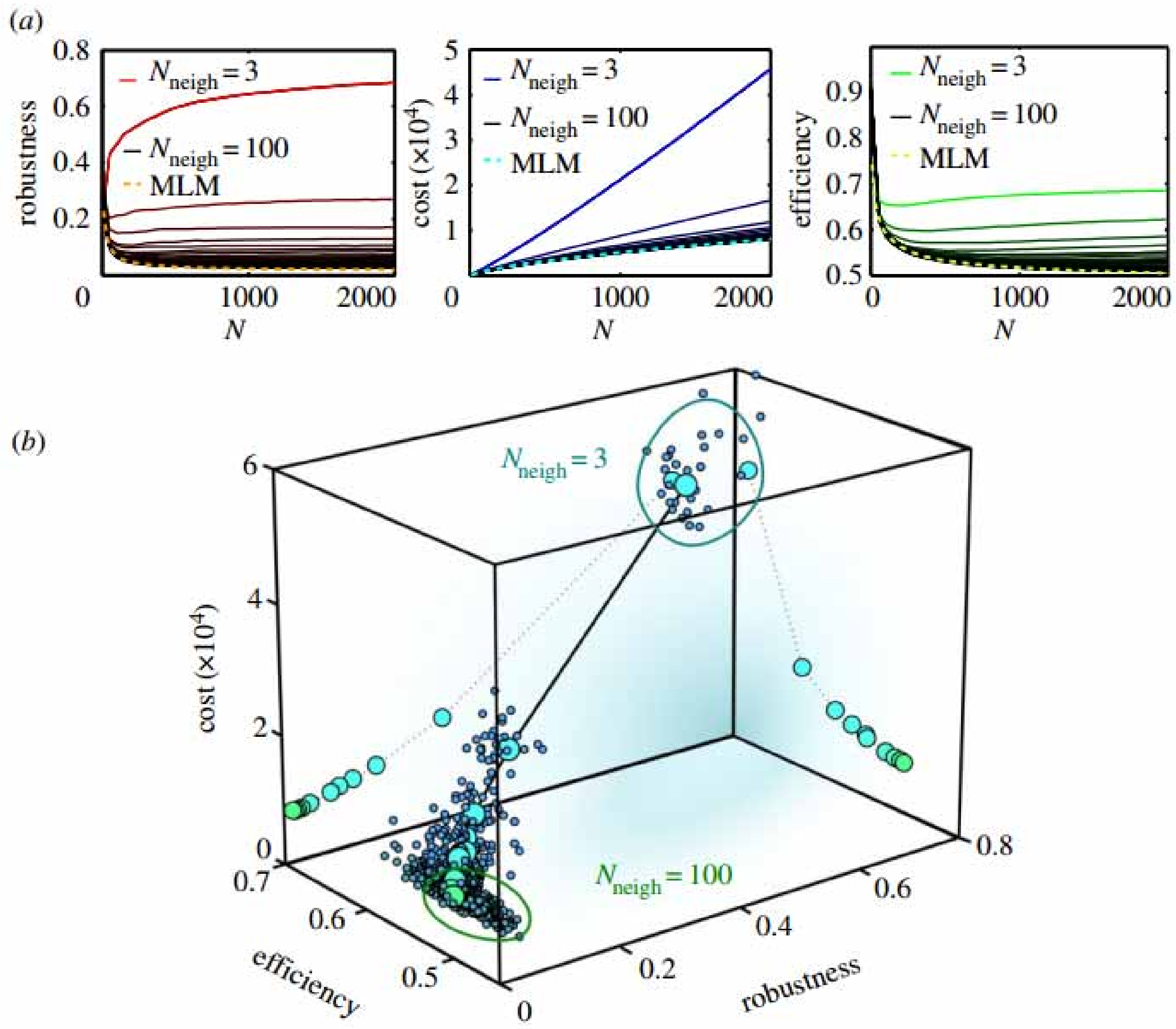
Reasons why ants find the shortest path (minimum linking model):

- 1) Earlier pheromones (the trail is completed earlier)
- 2) More pheromone (higher ant density)
- 3) Younger pheromone (less diffusion)

Soon, the ants will find the shortest path between their home and the food

Bottinelli, A., Van Wilgenburg, E., Sumpter, D. & Latty, T. 2015. Local cost minimization in ant transport networks: from small-scale data to large-scale trade-offs. Journal of The Royal Society Interface, 12, (112), 20150780, doi:10.1098/rsif.2015.0780.

No Free Lunch for the Ants



Bottinelli, A., Van Wilgenburg, E., Sumpter, D. & Latty, T. 2015. Local cost minimization in ant transport networks: from small-scale data to large-scale trade-offs. Journal of The Royal Society Interface, 12, (112), 20150780, doi:10.1098/rsif.2015.0780.

Real ants versus artificial ants (original paper 1996)

30 IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS–PART B: CYBERNETICS, VOL. 26, NO. 1, FEBRUARY 1996

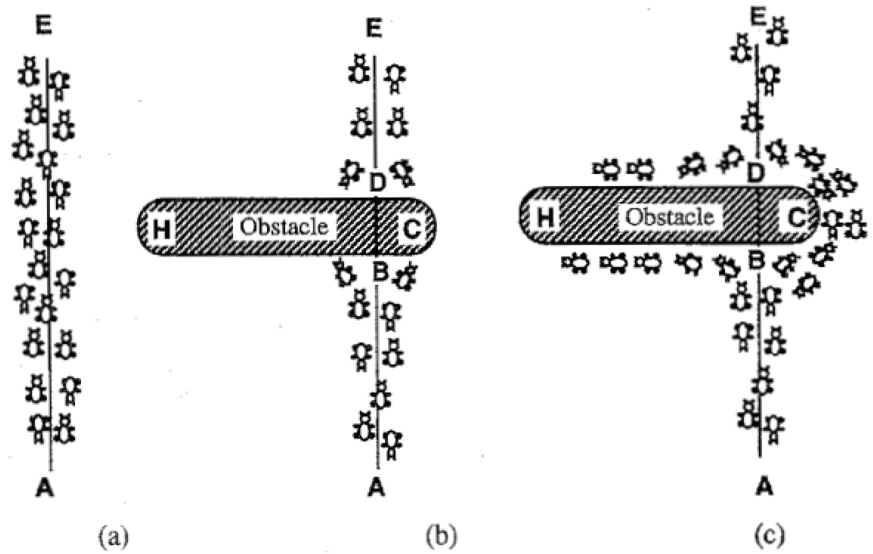


Fig. 1. An example with real ants. (a) Ants follow a path between points A and E. (b) An obstacle is interposed; ants can choose to go around it following one of the two different paths with equal probability. (c) On the shorter path more pheromone is laid down.

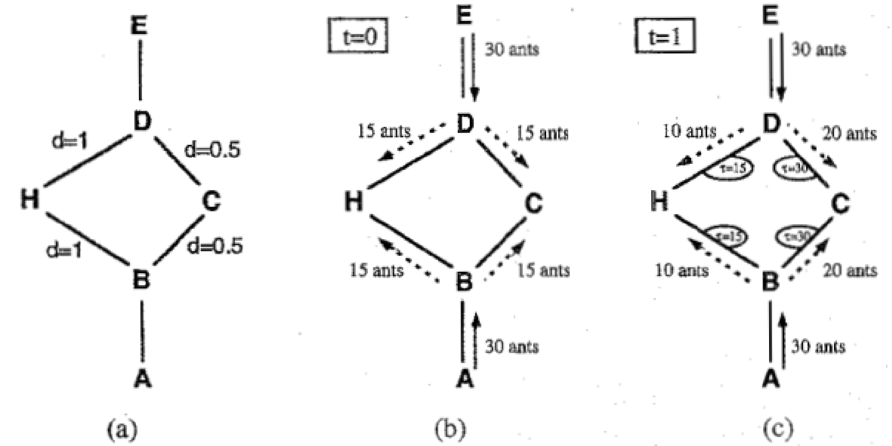


Fig. 2. An example with artificial ants. (a) The initial graph with distances. (b) At time $t = 0$ there is no trail on the graph edges; therefore ants choose whether to turn right or left with equal probability. (c) At time $t = 1$ trail is stronger on shorter edges, which are therefore, in the average, preferred by ants.

while walking an ant lays down at time t a pheromone trail of

Dorigo, M., Maniezzo, V. & Colorni, A. 1996. Ant system: optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 26, (1), 29-41, doi:10.1109/3477.484436.

ACO-Pseudocode

```
initialize pheromones τij
for each iteration do
  for k = 1 to number of ants do
     set out ant k at start node
     while ant k has not build a solution do
          choose the next node of the path
     enddo
  enddo
  update pheromones
enddo
return best solution found
```

ACO Basic Algorithm


```
initialize pheromones τij ;                  // usually identical, all τ0
place each ant k on a random city;
for each iteration do
  for i = 1 to number of ants do
      build a solution by applying (e-1) times:
            at city i, choose the next city j with
            probability given on next slide;
  end for                                    // e: number of edges of G
  eval the length of every solution build;
  if an improved solution is found
      then update the best solution;
  end if
  update pheromones (slides 11&12);
end for
return best solution found;
```

ACO Pseudo Code for Ant Colony System

Algorithm 1: Ant Colony Algorithm

```
Input  : ProblemSize, Population_size, m, ρ, β, σ, q0
Output: P_best
P_best ← CreateHeuristicSolution(ProblemSize);
Pbest_cost ← Cost(S_h);
Pheromone_init ← 1.0 / (ProblemSize × Pbest_const);
Pheromone ← InitializePheromone(Pheromone_init);
while ¬StopCondition() do
    for i = 1 to m do
        S_i ← ConstructSolution(Pheromone, ProblemSize, β, q0);
        Si_cost ← Cost(S_i);
        if Si_cost ≤ Pbest_cost then
            Pbest_cost ← Si_cost;
            P_best ← S_i;
        end
        LocalUpdateAndDecayPheromone(Pheromone, S_i, Si_cost, σ);
    end
    GlobalUpdateAndDecayPheromone(Pheromone, P_best, Pbest_cost, ρ);
end
return P_best;
```

Brownlee, J. 2011. Clever algorithms: nature-inspired programming recipes, Jason Brownlee.

What is the probability for selecting a particular path?

$$p_{ij} = \frac{[\tau_{ij}]^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\sum_{l \in J_i^k} [\tau(t)]^{\alpha} \cdot [\eta]^{\beta}}$$

- p_{ij} ... **probability** of ants that they, at a particular node i, select the route from node $i \rightarrow j$ (**“heuristic desirability”**)
- $\alpha > 0$ and $\beta > 0$ ... the **influence parameters** (α ... history coefficient, β ...heuristic coefficient) usually $\alpha \approx \beta \approx 2 < 5$
- τ_{ij} ... the **pheromone value** for the components, i.e. the amount of pheromone on edge (i, j)
- k ... the set of usable components
- J_i ... the set of nodes that ant k can reach from v_i (tabu list)
- $\eta_{ij} = \frac{1}{dij}$ ... attractiveness computed by a heuristic, indicating the “a-priori **desirability”** of the move

The pheromone on each edge is updated as:

$$\tau_{ij} = (1-\rho) \cdot \tau_{ij} + \Delta\tau_{ij}$$

With:

- ρ : the evaporation rate of the 'old' pheromone
- $\Delta\tau_{ij}$: the 'new' pheromone that is deposited by all ants on edge (i,j) calculated as:

$$\Delta\tau_{ij} = \sum_{k=0}^{m} \Delta\tau_{ij}^{k}$$

The pheromone that is deposited on edge (i,j) by ant k is calculated as:

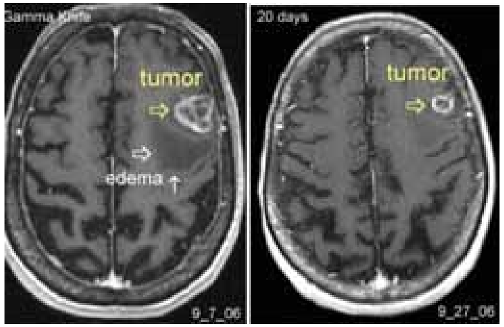
$$\Delta\tau_{ij}^{k} = \begin{cases} Q / L_k & if\ (i,j) \in T_k \\ 0 & otherwise \end{cases}$$

With:

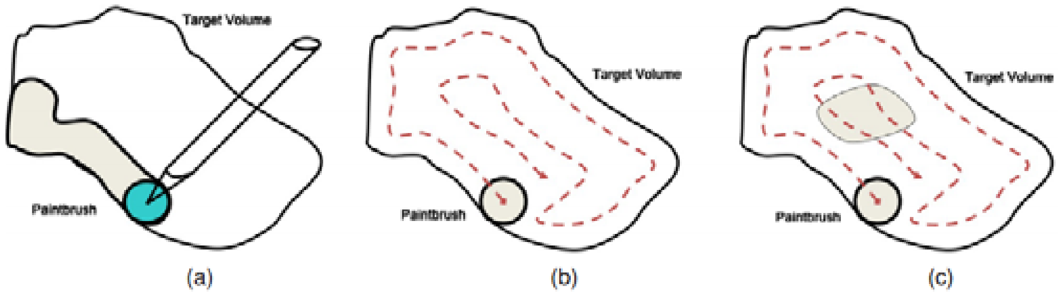
- Q : a heuristic parameter
- T_k : the path traversed by ant k
- L_k : the length of T_k calculated as the sum of the lengths of all the edges of T_k

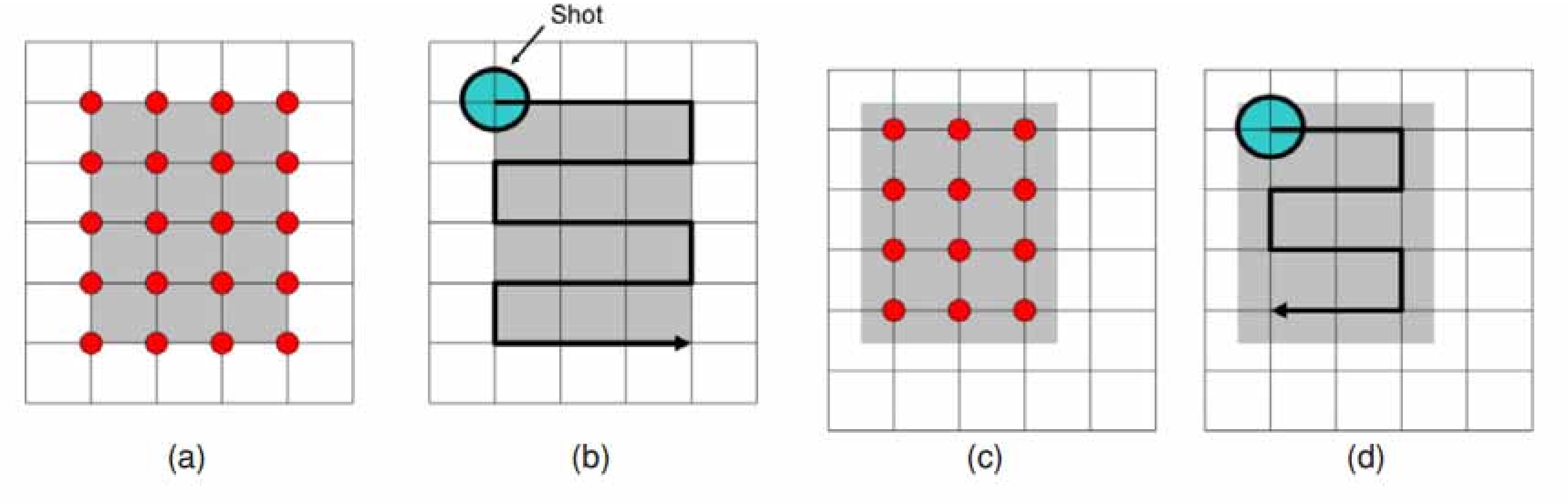
- The attractiveness η_{ij} of edge (i, j) is computed by a heuristic, indicating the a-priori desirability of that particular move
- The pheromone trail level τ_{ij} of edge (i, j) indicates how proficient it was in the past
- $\alpha = 0$ is a greedy approach and $\beta = 0$ represents the selection of tours that may not be optimal
- Consequently, we speak of a “trade-off” between speed and quality

Excursus: Traveling Salesman Problem = hard



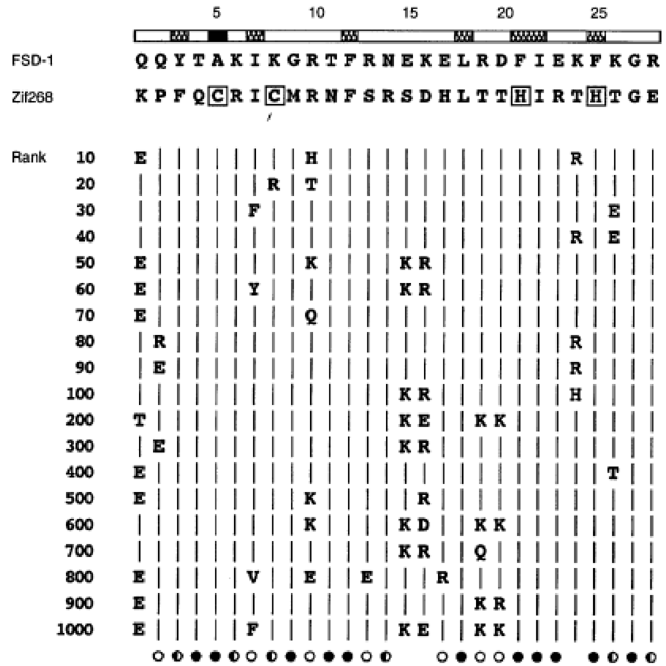
http://www.aboutcancer.com

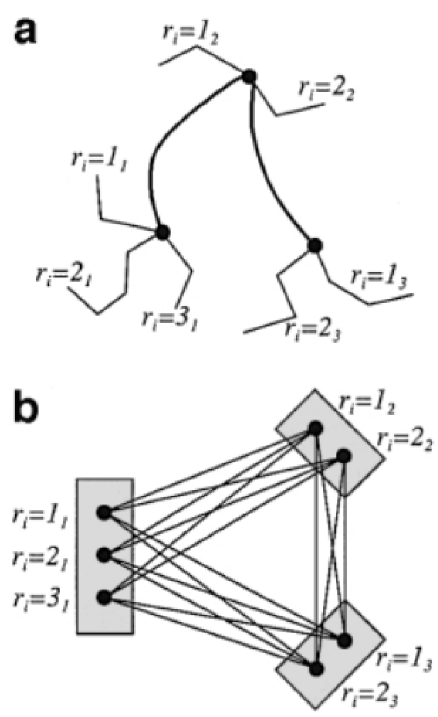




Luan, S. A., Swanson, N., Chen, Z. & Ma, L. J. 2009. Dynamic gamma knife radiosurgery. Physics in Medicine and Biology, 54, (6), 1579-1591, doi:10.1088/0031-9155/54/6/012.

Protein Design is a hard problem

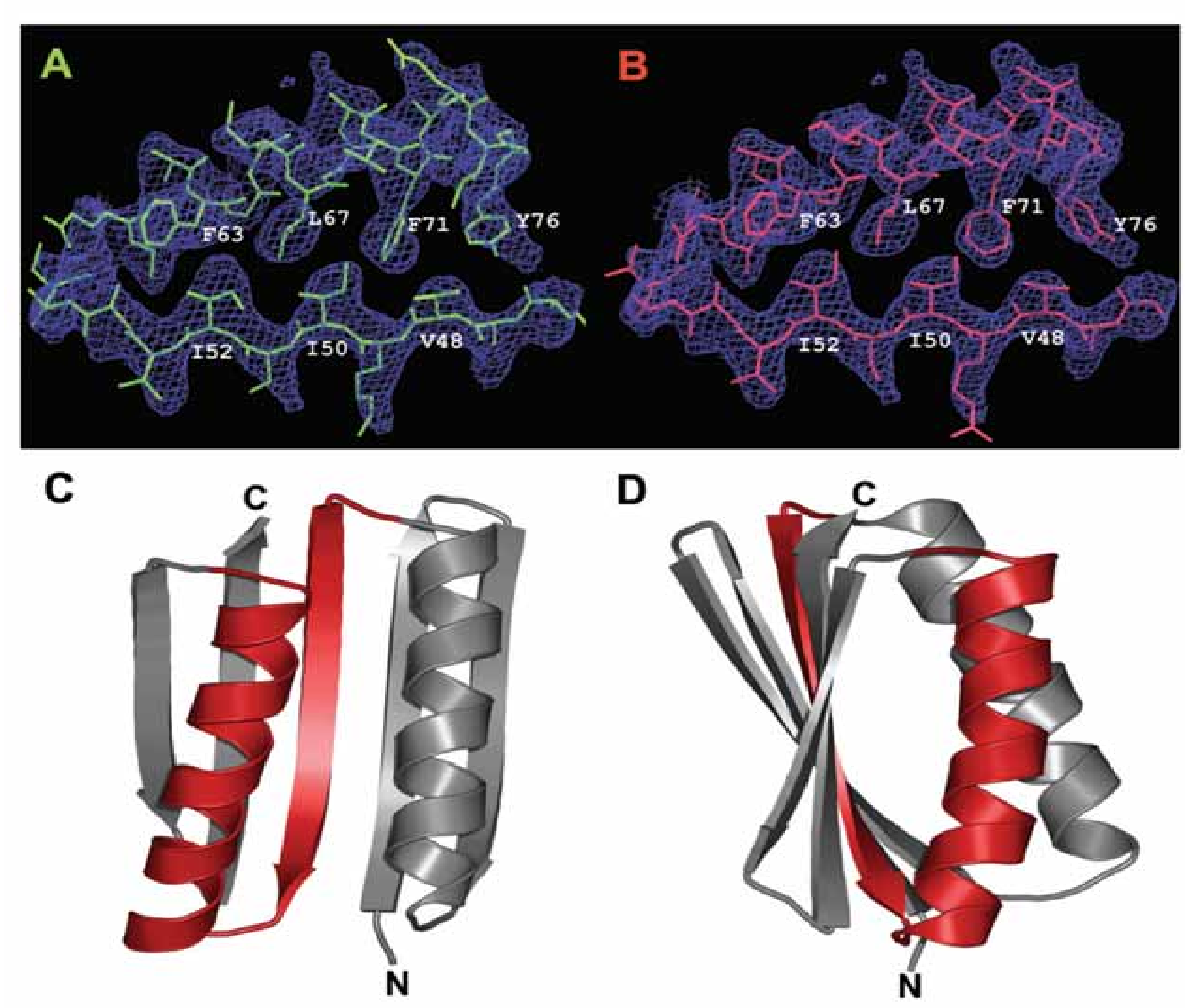




Pierce, N. A. & Winfree, E. 2002. Protein design is NP-hard. Protein Engineering, 15, (10), 779-782.

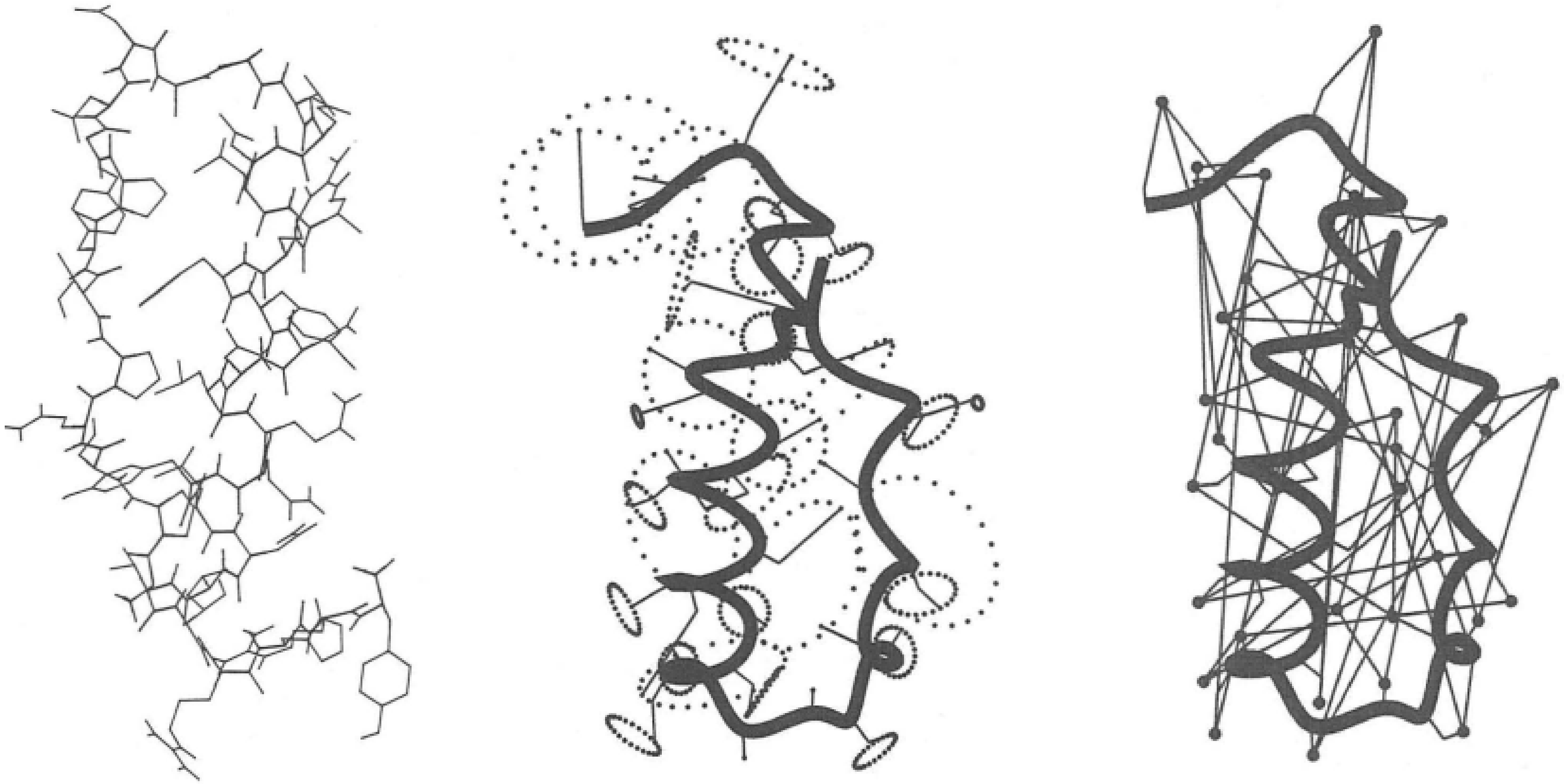
Dahiyat, B. I. & Mayo, S. L. 1997. De novo protein design: fully automated sequence selection. *Science*, 278, (5335), 82-87.

Protein Design is a big challenge and is important for PM



Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. 2003. Design of a novel globular protein fold with atomic-level accuracy. Science, 302, (5649), 1364-1368.

Protein Folding is a TSP



Bohr, H. & Brunak, S. 1989. A travelling salesman approach to protein conformation. Complex Systems, 3, 9-28

- Desirability $\eta_{ij} = \frac{1}{d_{ij}}$
- The tabu-list contains all places (="cities) an ant has visited already.
- $N = e$
- Adding "elitary ant" with
- $\alpha = 1, \beta = 5, \varrho = 0.5, Q = 100, t_0 = 10^{-6}, b = 5$

$$\tau_{ij} = (1-\rho)\cdot\tau_{ij} + \Delta\tau_{ij} + b\Delta\tau_{ij}^{best}$$

$$\Delta\tau_{ij}^{best} = \begin{cases} Q / L_{best} & if\ (i,j) \in best \\ 0 & otherwise \end{cases}$$

Advantages:

- Applicable to a broad range of optimization problems.
- Can be used in dynamic applications (adapts to changes such as new distances, etc.).
- Can compete with other global optimization techniques like genetic algorithms and simulated annealing.

Disadvantages:

- Only applicable for discrete problems.
- Theoretical analysis is difficult.

I. Represent the problem in the form of a weighted graph, on which ants can build solutions
II. Define the meaning of the pheromone trails
III. Define the heuristic preference for the ant while constructing a solution
IV. Choose a specific ACO algorithm and apply to the problem being solved
V. Tune the parameters of the ACO algorithm

Example Applications of ACO

- Scheduling
- Routing problems
- – Traveling Salesman Problem (TSP)
- – Vehicle routing
- – Network routing
- Set-problems
- – Multi-Knapsack
- – Max Independent Set
- – Set Covering
- Many others, e.g.
- – Shortest Common Sequence
- – Constraint Satisfaction
- – 2D-HP protein folding
- – Edge detection

Biology (Ant Foraging)	ACO Algorithm
Ant	Individual (agent) used to build (construct) a solution
Ant Colony	Population (colony) of cooperating individuals
Pheromone Trail	Modification of the environment caused by the artificial ants in order to provide an indirect mean of communication with other ants of the colony. Allows assessment of the quality of a given edge on a graph.
Pheromone Evaporation	Reduction in the pheromone level of a given path due to aging.

Example: ACO in health informatics

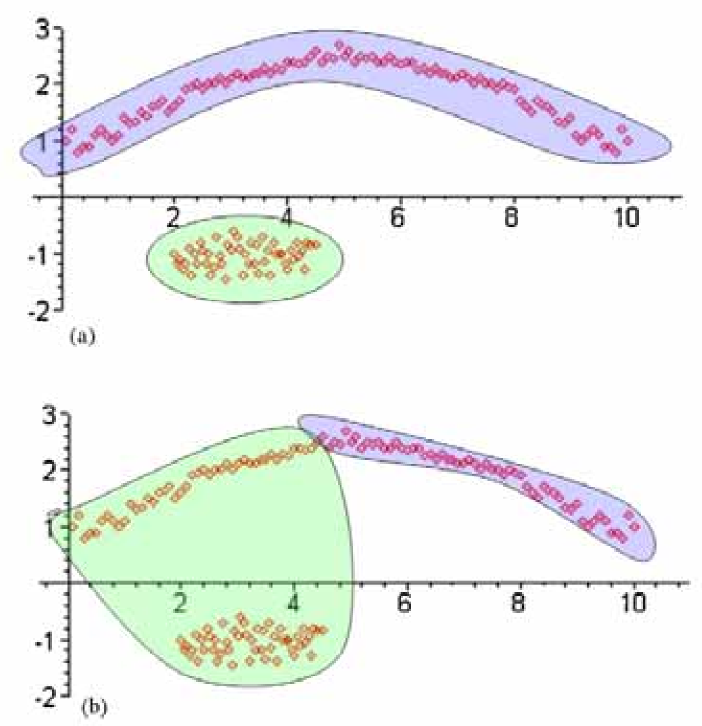


Fig. 1. Data partition examples: (a) intuitive partition; (b) ACO partition.

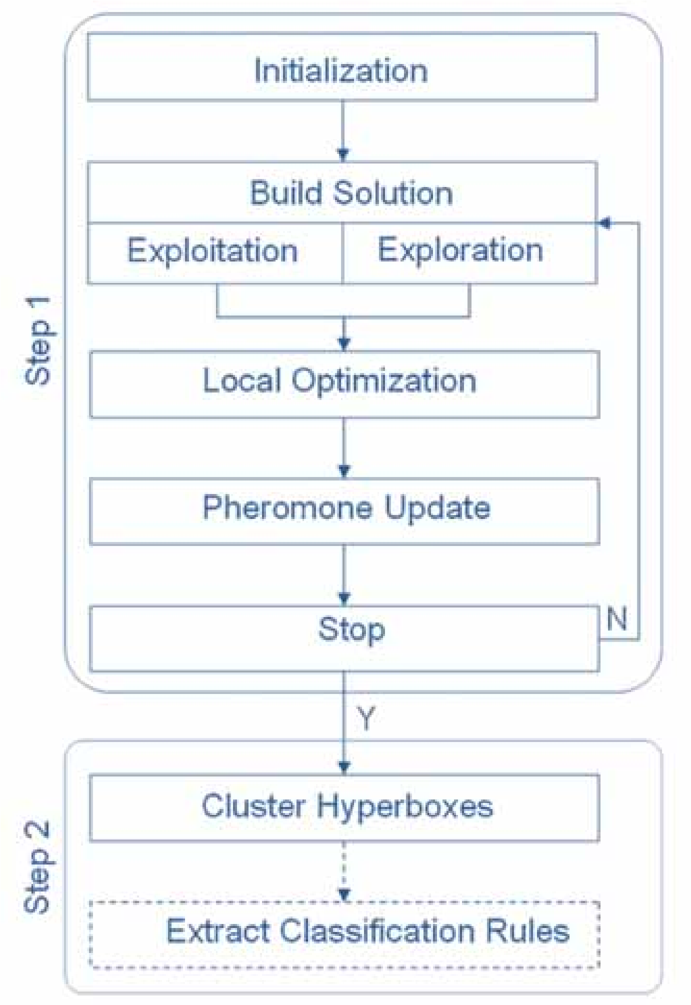
$$H = \{H_1, H_2, \ldots, H_l, \ldots, H_C\},$$
$$H_l \subset \Re^n,$$
$$y = \{y_1, y_2, \ldots, y_j, \ldots, y_n\},$$
$$y \in H_l \Rightarrow a_{lj} \le y_j \le b_{lj}, \quad a_l, b_l \in \Re^n$$

$$X = \{x_1, x_2, \ldots, x_i, \ldots, x_N\} \subset \Re^n,$$
$$D = \{D_1, D_2, \ldots, D_j, \ldots, D_n\},$$
$$\forall y \in \Re^n,$$
$$y \in H_l \Rightarrow x_{ij} - \frac{D_j}{2} \le y_j \le x_{ij} + \frac{D_j}{2},$$

A hyperbox defines a region in an n-dimensional space and is fully described by two vectors, usually its two extreme points: a_i which is the lower bound and b_i, the upper bound.

Ramos, G. N., Hatakeyama, Y., Dong, F. & Hirota, K. 2009. Hyperbox clustering with Ant Colony Optimization (HACO) method and its application to medical risk profile recognition. Applied Soft Computing, 9, (2), 632-640, doi:10.1016/j.asoc.2008.09.004.

Example: ACO in health informatics



$$C = rd\left(\alpha \cdot \prod_{j=1}^{n} \frac{||\max(x_{ij}) - \min(x_{ij})||}{D_j}\right)$$

$$S_r = \{H_{r1}, H_{r2}, \ldots, H_{rC}\},$$

$$d_r = \frac{1}{N}\sum_{i=1}^{N} f_r(x_i),$$

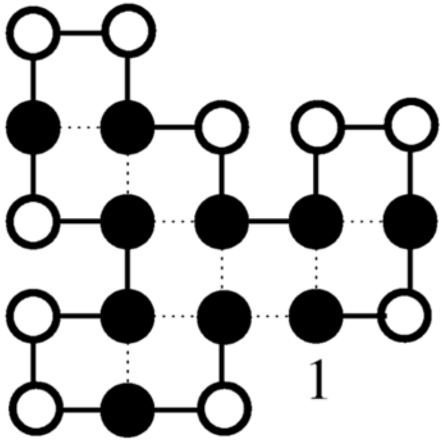
$$x_i \in X,$$
$$l, m \in \{1, 2, \ldots, C\}, \quad l < m,$$
$$f_r(x_i) = \begin{cases} 1, & x_i \in H_{rm}, x_i \notin H_{rl}, \\ 0, & \text{otherwise.} \end{cases}$$

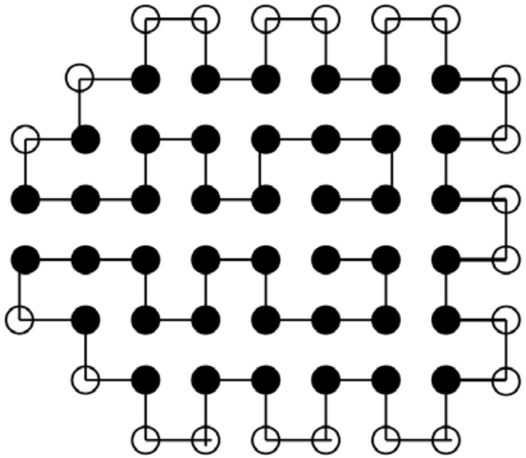
$$\forall\, l, m \in \{1, 2, \ldots, C\}, \quad l \neq m \Rightarrow S_{rl} \neq S_{rm}.$$

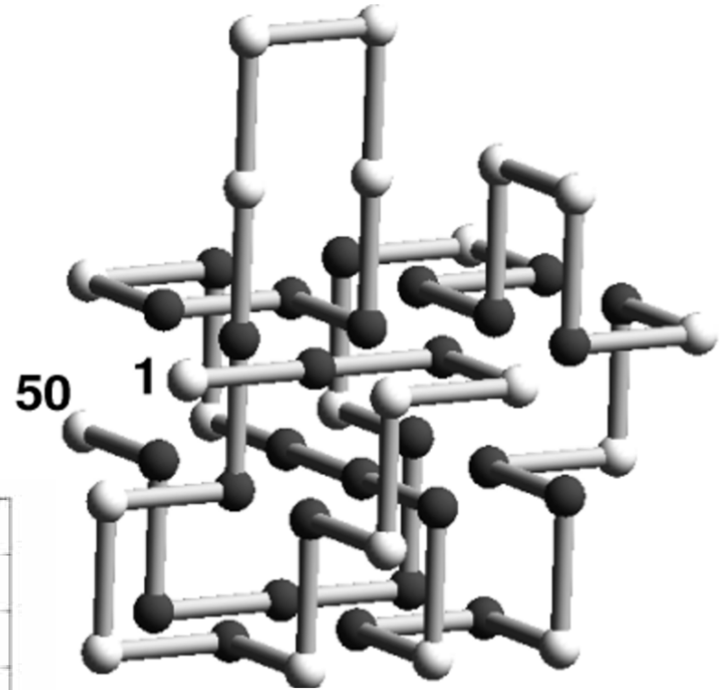
$$p_{il} = \frac{\tau_{il}}{\sum_{i=1}^{N} \tau_{il}}$$

Ramos, G. N., Hatakeyama, Y., Dong, F. & Hirota, K. 2009. Hyperbox clustering with Ant Colony Optimization (HACO) method and its application to medical risk profile recognition. Applied Soft Computing, 9, (2), 632-640, doi:10.1016/j.asoc.2008.09.004.

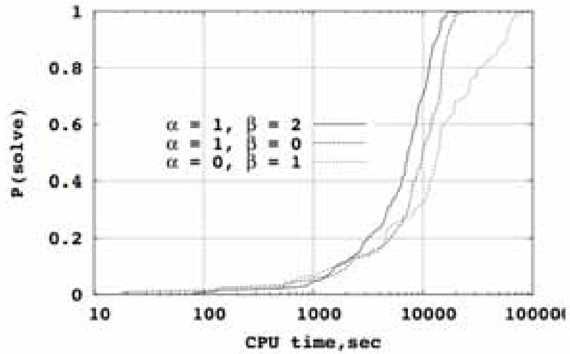
Example: HPHPPHHPHPPHPHHPPHPH 2D HP model



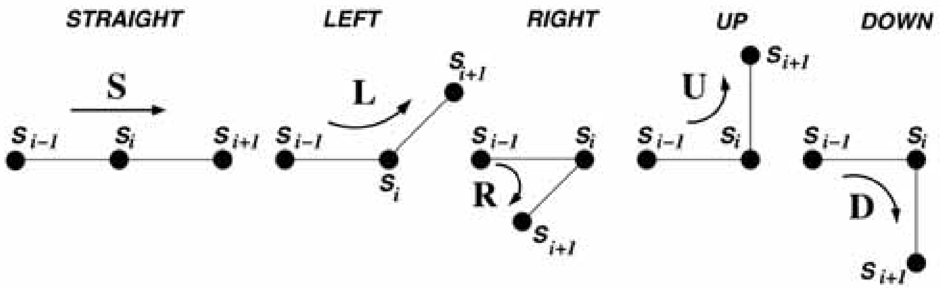





```
procedure ACO
    initialise pheromone trails;
    while (termination condition not satisfied) do
        construct candidate conformations;
        perform local search;
        update pheromone values;
    end
end
```



$$p_{i,d} := \frac{[\tau_{i,d}]^{\alpha}[\eta_{i,d}]^{\beta}}{\sum_{e\in\{S,L,R,U,D\}}[\tau_{i,e}]^{\alpha}[\eta_{i,e}]^{\beta}}$$



ID	E	HZ	CHCC	CG
S2-1	**-32**	-31(4 hrs)	**-32** (30 min)	**-32** (9.4 min)
S2-2	**-34**	-32 (18 hrs)	**-34** (2.3 min)	**-34** (35 min)
S2-3	**-34**	-31 (23 hrs)	**-34** (30 min)	**-34** (62 min)
S2-4	**-33**	-30 (19 days)	**-33** (71 min)	**-33** (29 min)
S2-5	**-32**	-30 (1.3 days)	**-32** (32 min)	**-32** (12 min)
S2-6	**-32**	-29 (2.1 days)	**-32** (80 min)	**-32** (460 min)
S2-7	**-32**	-29 (2.5 days)	**-32** (110 min)	**-32** (64 min)
S2-8	**-31**	-29 (4 hrs)	**-31** (530 min)	**-31** (38 min)
S2-9	**-34**	-31(4.5 hrs)	**-34** (8.3 min)	-33
S2-10	**-33**	**-33** (1.1 hr)	**-33** (4.8 min)	**-33**(1.1 min)

Shmygelska, A. & Hoos, H. H. 2005. An ant colony optimisation algorithm for the 2D and 3D hydrophobic polar protein folding problem. BMC bioinformatics, 6, (1), 1.

Digression: Simulated Annealing

- Simulated annealing presents an optimization technique that can:
- (a) process cost functions possessing quite arbitrary degrees of nonlinearities, discontinuities,
- and stochasticity;
- (b) process quite arbitrary boundary conditions and constraints imposed on these cost func-
- tions;
- (c) be implemented quite easily with the degree of coding quite minimal relative to other
- nonlinear optimization algorithms;
- (d) statistically guarantee finding an optimal solution

Ingber, L. 1993. Simulated annealing: Practice versus theory. Mathematical and computer modelling, 18, (11), 29-57.

04 Ant’s and Collective Intelligence Human-in-the-loop

http://functionlearning.com

Demos of experiments for *The Human Kernel*

This page contains links to the experiments described in Section 4 of The Human Kernel [bibtex].

- **Part 1: Extrapolating from smooth functions.**
  In the first experiment, described in Section 4.2 of the paper, participants were asked to exptrapolate from several functions, where the true underlying relationships were draws from a Gaussian process with a rational quadratic kernel. [Link]
- **Part 2: Extrapolating from smooth functions.**
  In the second experiment, described in Section 4.3 of the paper, participants were asked to exptrapolate from functions that are difficult or impossible for conventional Gaussian process methods to capture. [Link]
- **Part 3: Preference for smoothness/simplicity.**
  In the third experiment, described in Section 4.4 of the paper, participants were asked to express their preferences over different kinds of explanations or underlying relationships, given a small number of data points. [Link]

Wilson, A. G., Dann, C., Lucas, C. & Xing, E. P. The Human Kernel. In: Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R., eds. Advances in Neural Information Processing Systems, NIPS 2015, 2015 Montreal. 2836-2844.

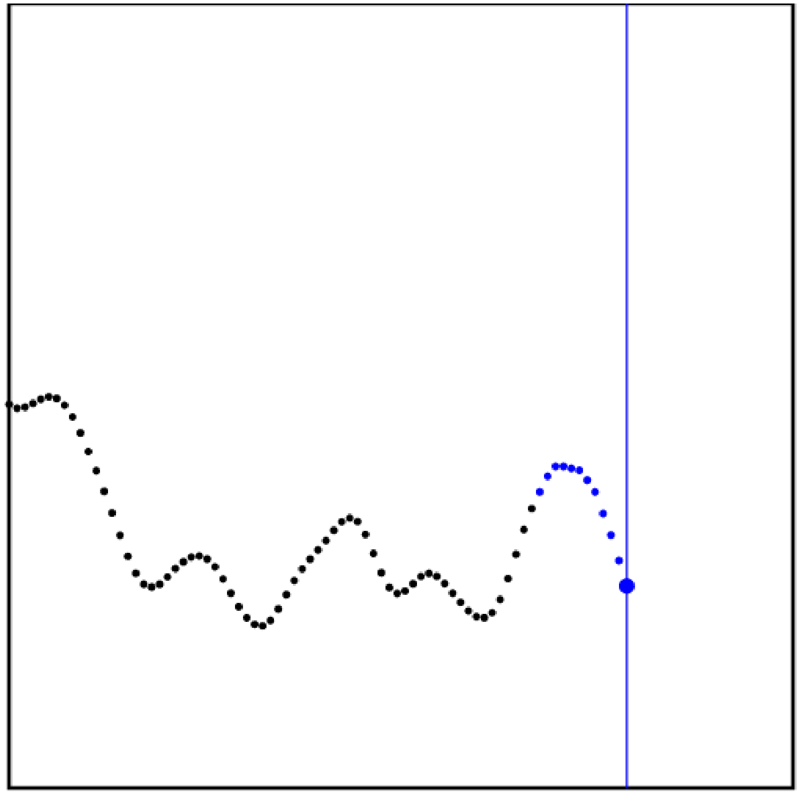
The Human Kernel Experiment (2/3)

Judgment 12 out of 33

This is the first function from the system. Please try to predict the new points as well as y

Please click along the blue line to say what you think the height of the point is for that lo

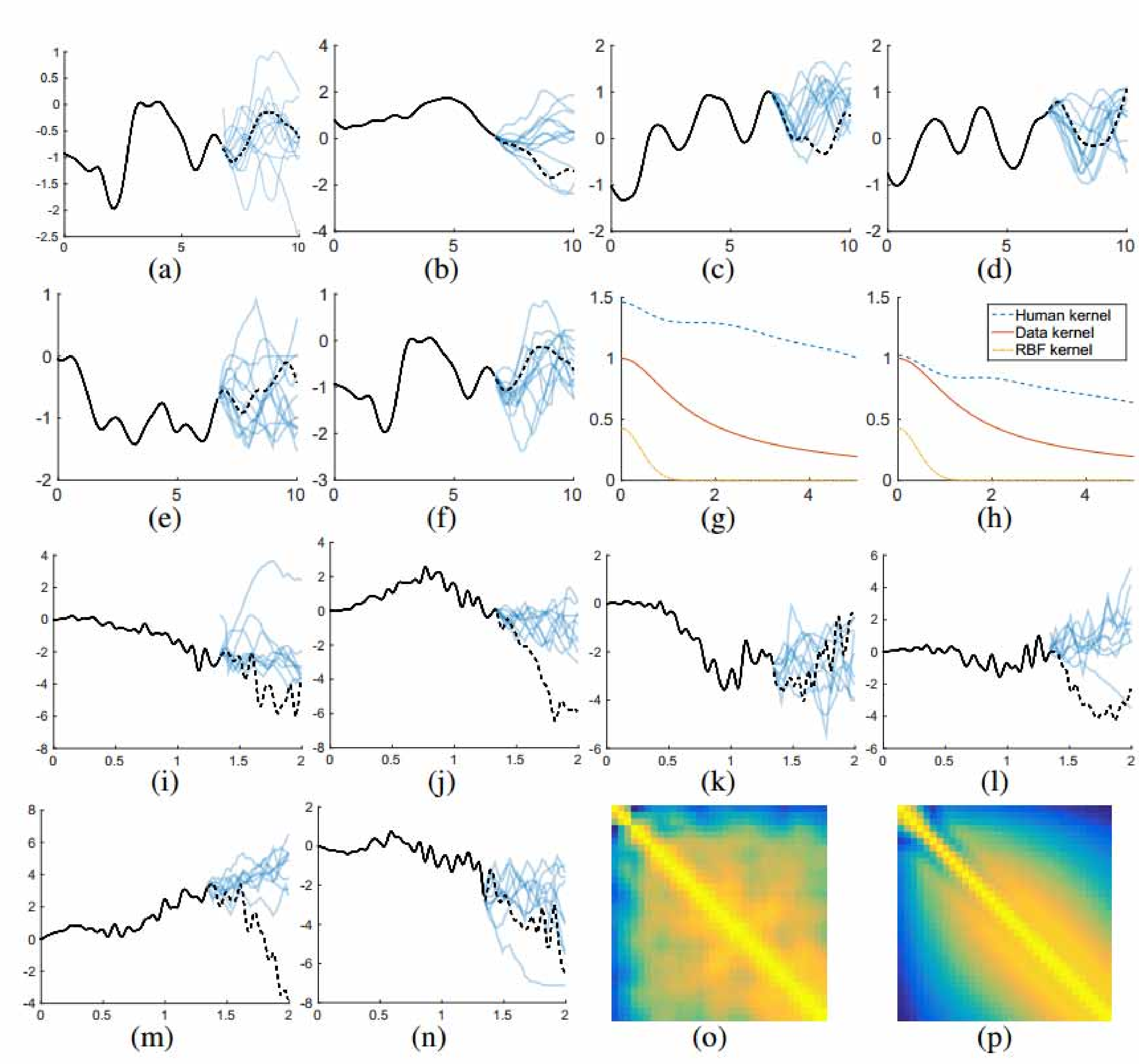
Once you have selected a position along the line, **hit the 's' key to submit the point**.



Wilson, A. G., Dann, C., Lucas, C. & Xing, E. P. The Human Kernel. In: Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R., eds. Advances in Neural Information Processing Systems, NIPS 2015, 2015 Montreal. 2836-2844.

The Human Kernel Experiment (3/3)

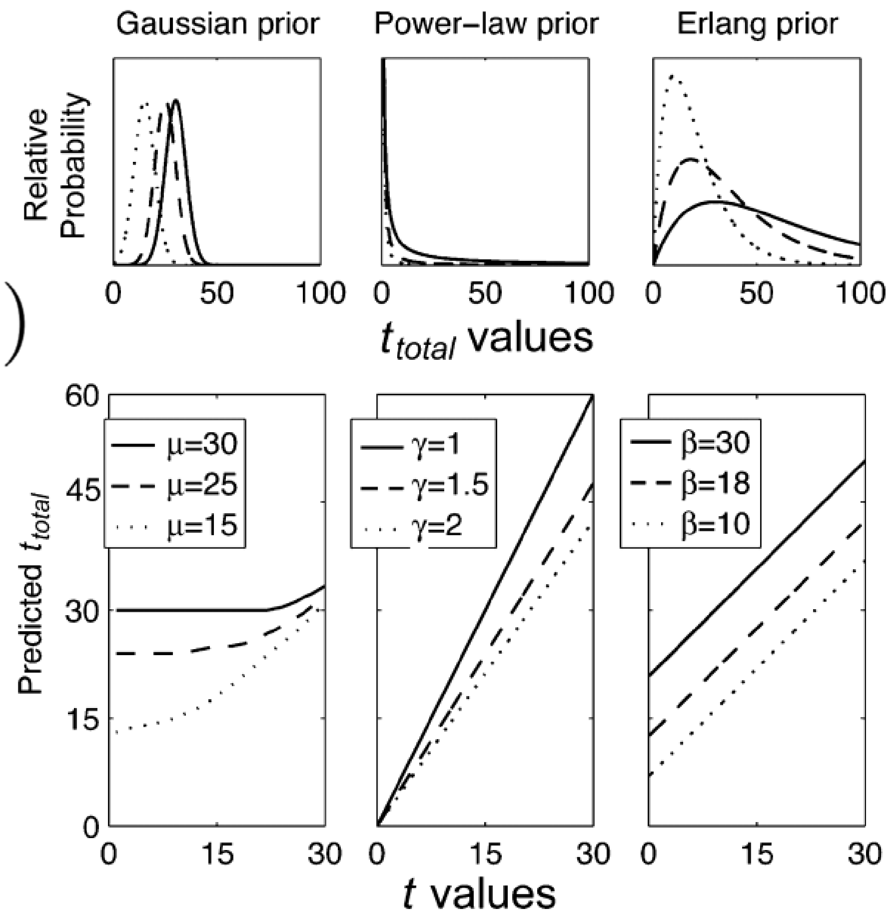
Wilson, A. G., Dann, C., Lucas, C. & Xing, E. P. The Human Kernel. In: Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R., eds. Advances in Neural Information Processing Systems, NIPS 2015, 2015 Montreal. 2836-2844.



Bayesian predictor computes a probability distribution

Life spans: Insurance agencies employ actuaries to make predictions about people's life spans—the age at which they will die—based upon demographic information. If you were assessing an insurance case for an 18-year-old man, what would you predict for his life span?

$$p(t_{\text{total}}|t) \propto p(t|t_{\text{total}})p(t_{\text{total}})$$



Griffiths, T. L. & Tenenbaum, J. B. 2006. Optimal predictions in everyday cognition.
Psychological science, 17, (9), 767-773, doi:10.1111/j.1467-9280.2006.01780.x.

Remember: How do grow a mind

Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. 2011. How to grow a mind: Statistics, structure, and abstraction. Science, 331, (6022), 1279-1285, doi:10.1126/science.1192788.

Ants communicate via non-linear pheromones trails

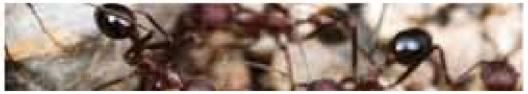


Figure 2. Pharaoh's ants, *Monomorium pharaonis*, form branching networks of pheromone trails.

Here the network has been formed on a smoked glass surface to aid visualisation. (Image courtesy of Duncan Jackson.)

Sumpter, D. J. T. & Beekman, M. 2003. From nonlinearity to optimality: pheromone trail foraging by ants. Animal Behaviour, 66, (2), 273-280, doi:10.1006/anbe.2003.2224.

Problem Solving: Humans vs. Computers

When is the human *) better?

***) human intelligence/natural intelligence/human mind/human brain/human learning**

- **Natural Language Translation/Curation**

  Machine cannot understand the context of sentences [3]

- **Unstructured problem solving**

  Without a pre-set of rules, a machine has trouble solving the problem, because it lacks the creativity required for it [1]

- **NP-hard Problems**

  Processing times are exponential and makes it almost impossible to use machines for it, so human still stays better [4]

When is the computer **) better?

****) Computational intelligence, Artificial Intelligence/ Machine Learning algorithms**

- **High-dimensional data processing**

  Humans are very good at dimensions less or equal than 3, but computers can process data in arbitrarily high dimensions

- **Rule-Based environments**

  Difficulties for humans in rule-based environments often come from not recognizing the correct goal in order to select the correct procedure or set of rules [2]

- **Image optimization**

  Machine can look at each pixel and apply changes without human personal biases, and with more speed [1]

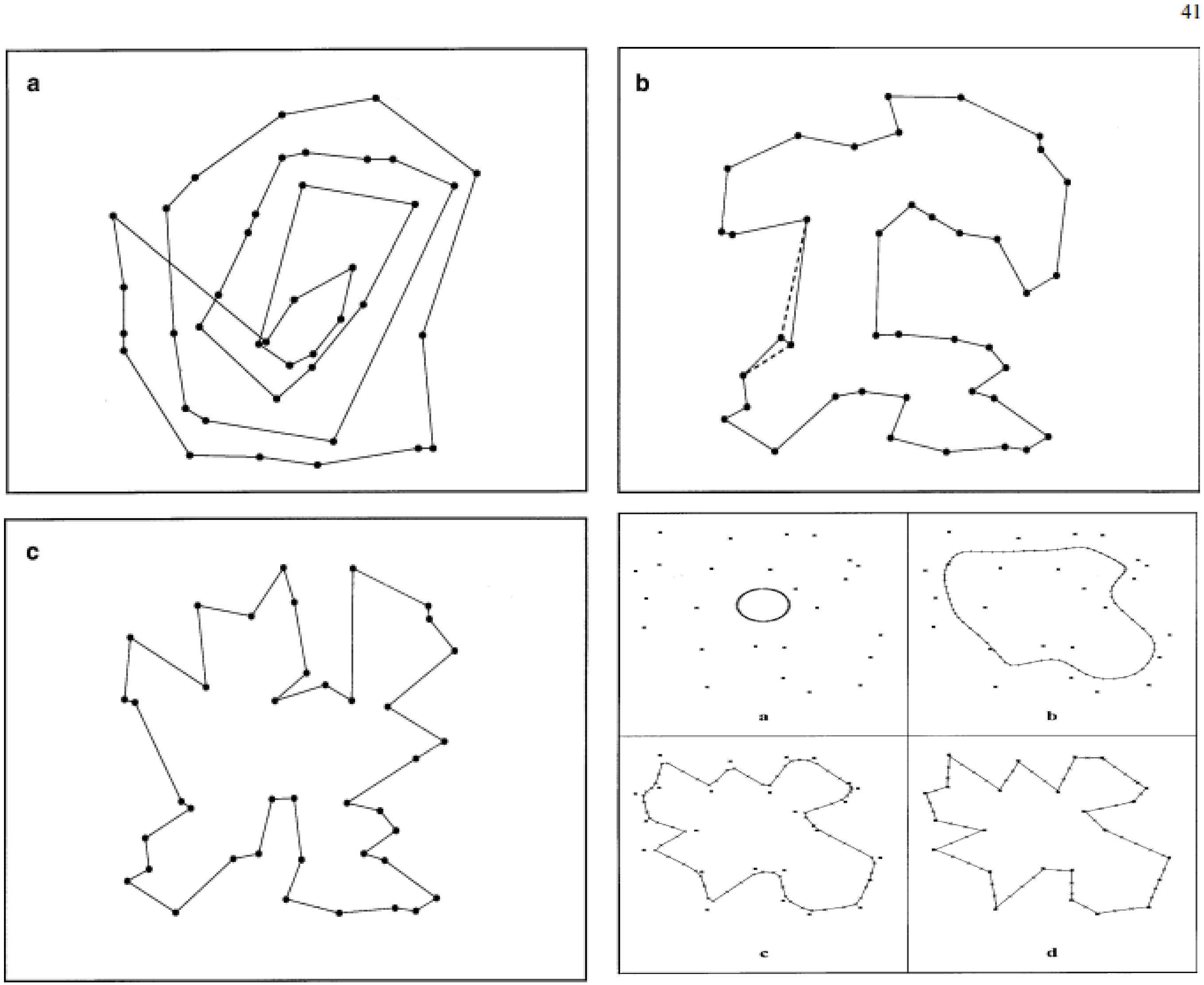
[1] https://www.instartlogic.com/blog/man-vs-machine-learning-based-optimizations

[2] Cummings, Mary Missy. "Man versus machine or man+ machine?." *Intelligent Systems, IEEE* 29.5 (2014): 62-69.

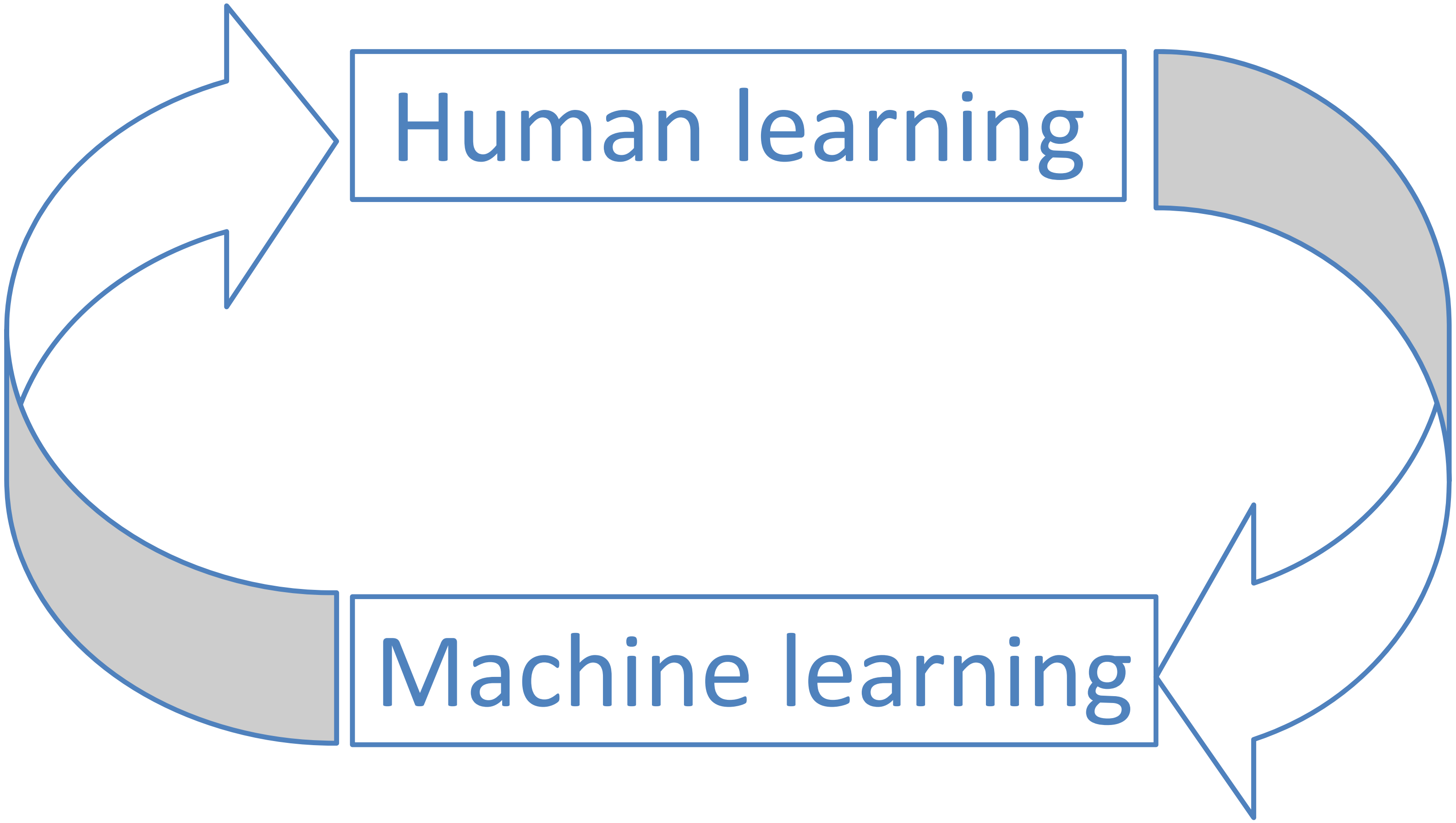
[3] Pizlo, Zygmunt, Anupam Joshi, and Scott M. Graham. "Problem Solving in Human Beings and Computers (formerly: Heuristic Problem Solving)." (1994).

[4] Griffiths, Thomas L. "Connecting human and machine learning via probabilistic models of cognition." *INTERSPEECH*. 2009.

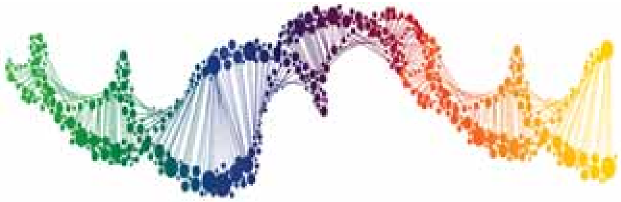
Human learning	Machine learning
Categorization	Density estimation
Causal learning	Graphical models
Function learning	Regression
Representations	Nonparametric Bayes
Language	Probabilistic grammars
Experiment design	Inference algorithms



Vickers, D., Butavicius, M., Lee, M. & Medvedev, A. 2001. Human performance on visually presented traveling salesman problems. Psychological Research, 65, (1), 34-45, doi:10.1007/s004260000031.


Human learning
Machine learning

- Drilling of circuit board
- Warehouse supply chain optimization
- Hospital Organization optimization
- Route planner
- DNA sequencing, Protein, etc.



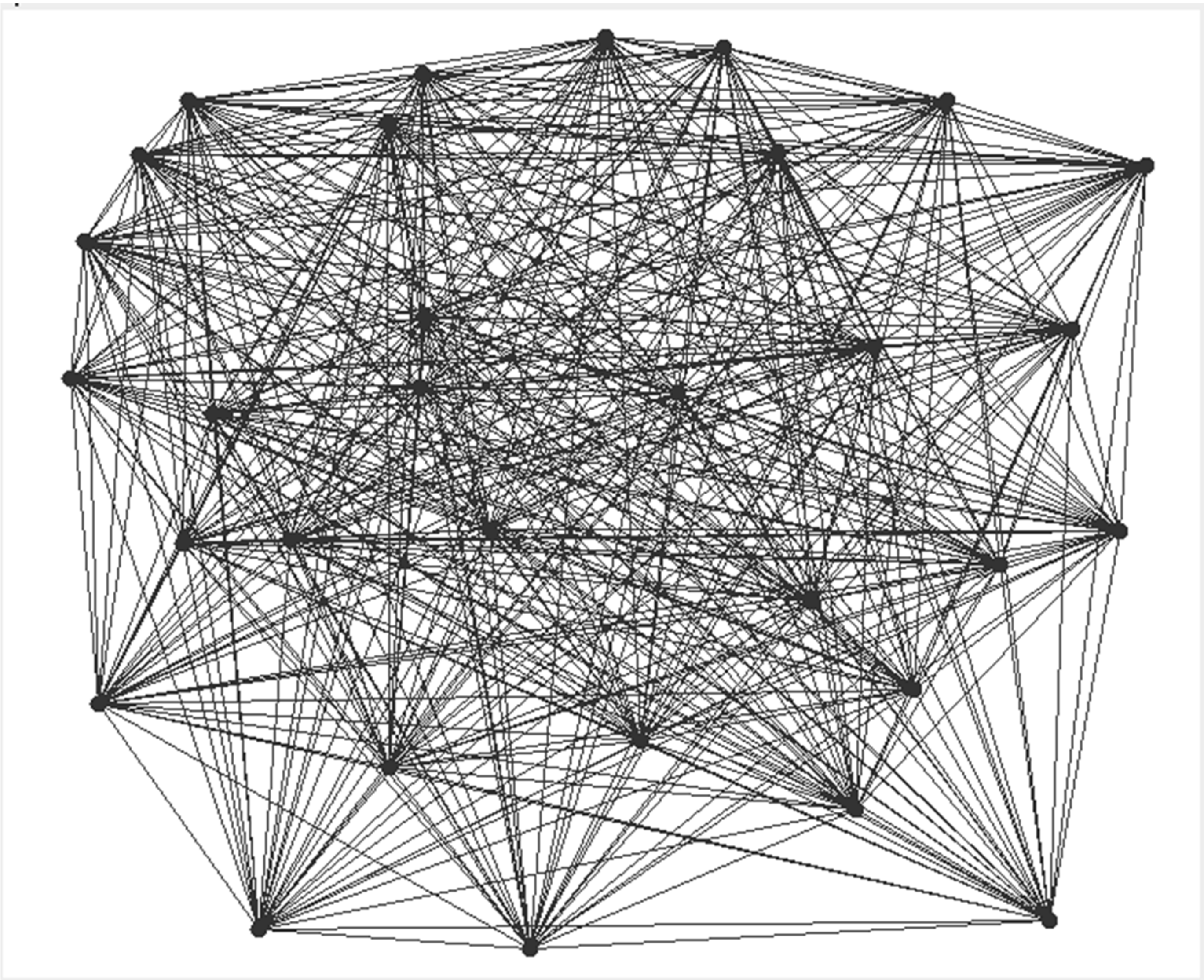
Dorigo, Marco, and Thomas Stützle. "Ant colony optimization: overview and recent advances." *Techreport, IRIDIA, Universite Libre de Bruxelles* (2009).

- Nature inspired Algorithm
- Swarm intelligence
- Artificial Ants
- Pheromone trail
- **Decision based on pheromones**

http://www.sciencemag.org/news/2015/12/bipolar-drug-turns-foraging-ants-scouts

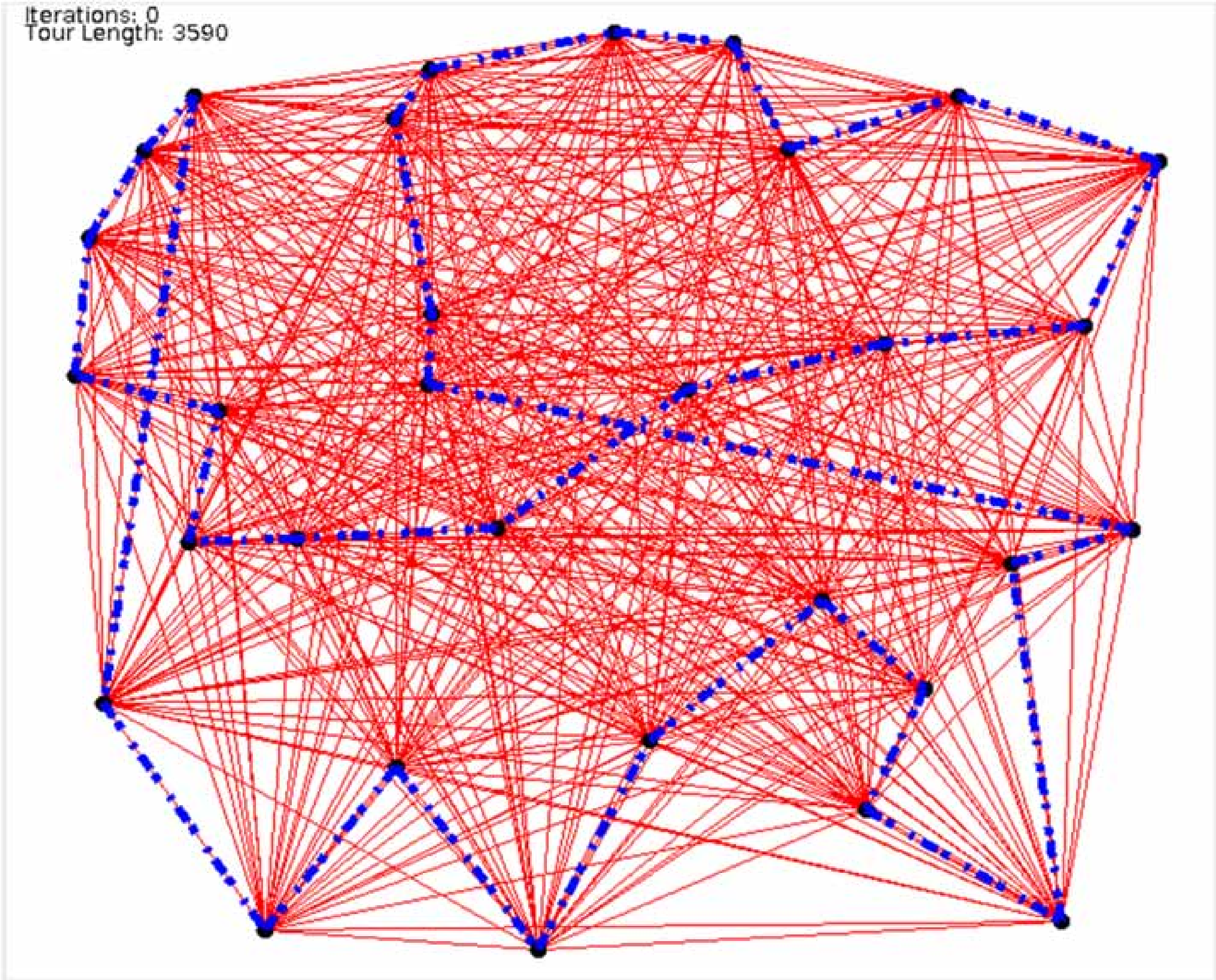
Dorigo, Marco, Mauro Birattari, and Thomas Stützle. "Ant colony optimization." *Computational Intelligence Magazine, IEEE* 1.4 (2006): 28-39.

Ant-Algorithm

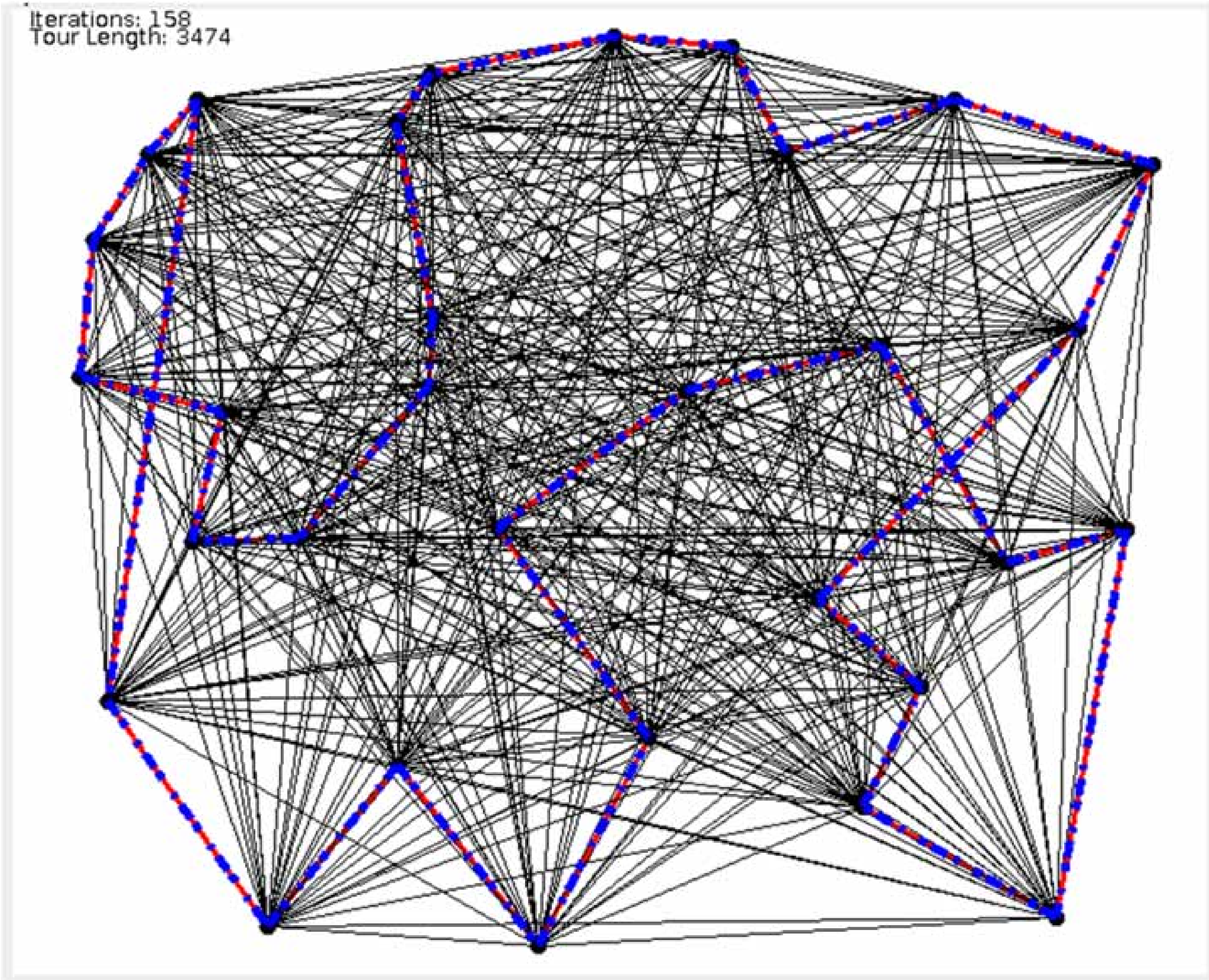


Source-Code: https://github.com/bogdan-ivanov/ants_aco

Initialisation:

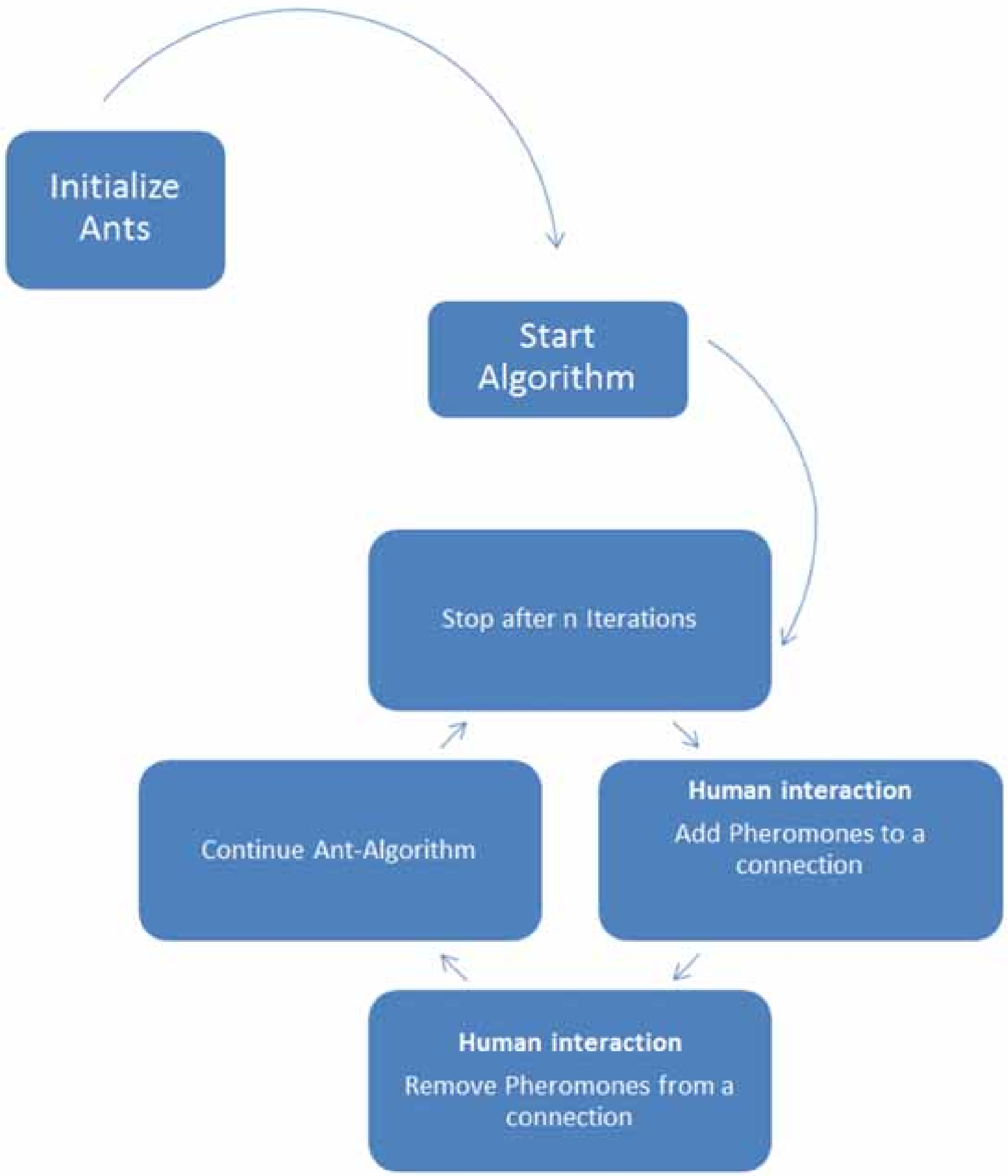


Result of Ant-Algorithm

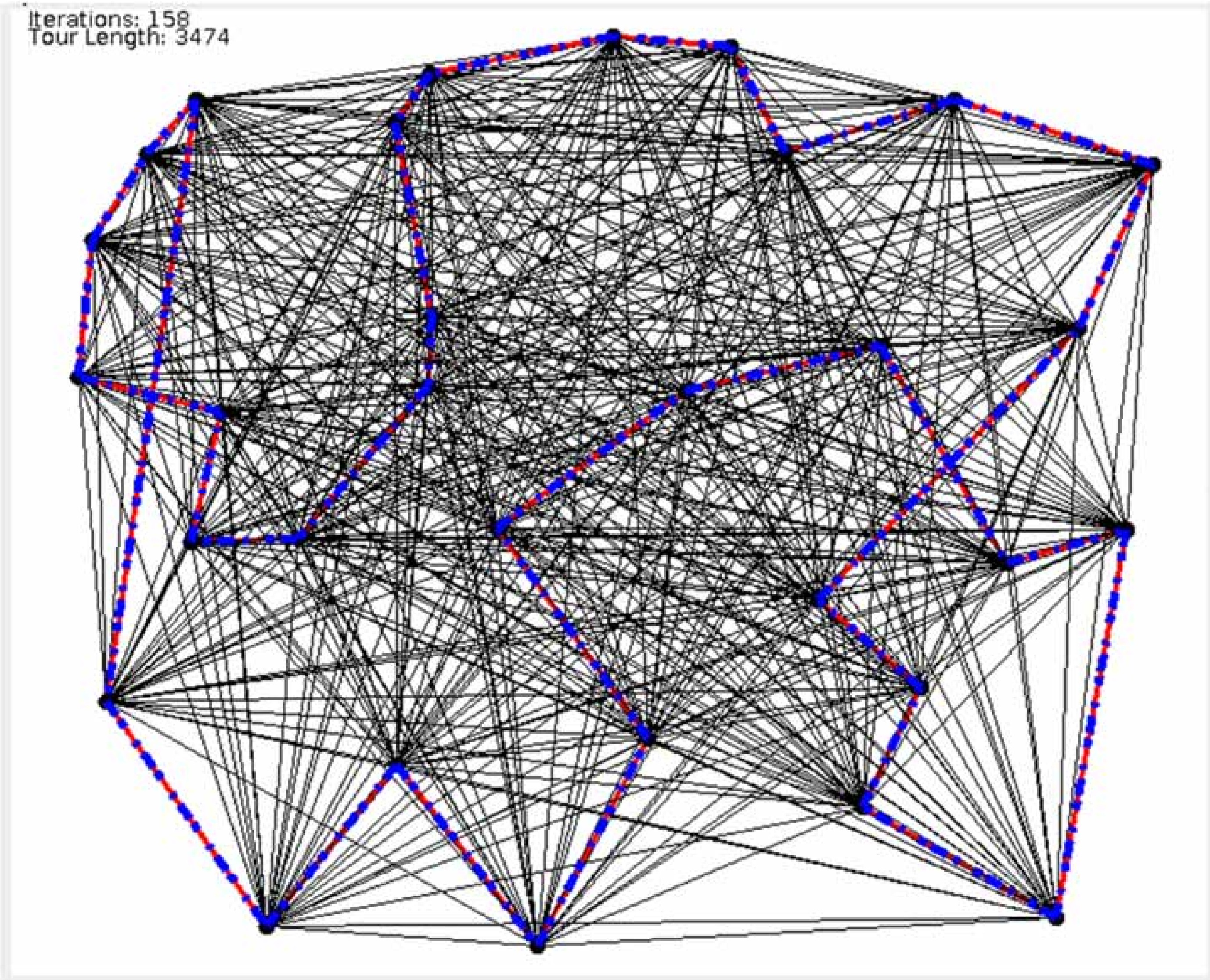


- What are the problems with the Ant-Algorithm?
  - Wrong Initialization
- What is the benefit of the interaction? How to measure the benefit?
  - Reduce of length
- When is an interaction with the Human possible?
  - Change the ant’s behavior

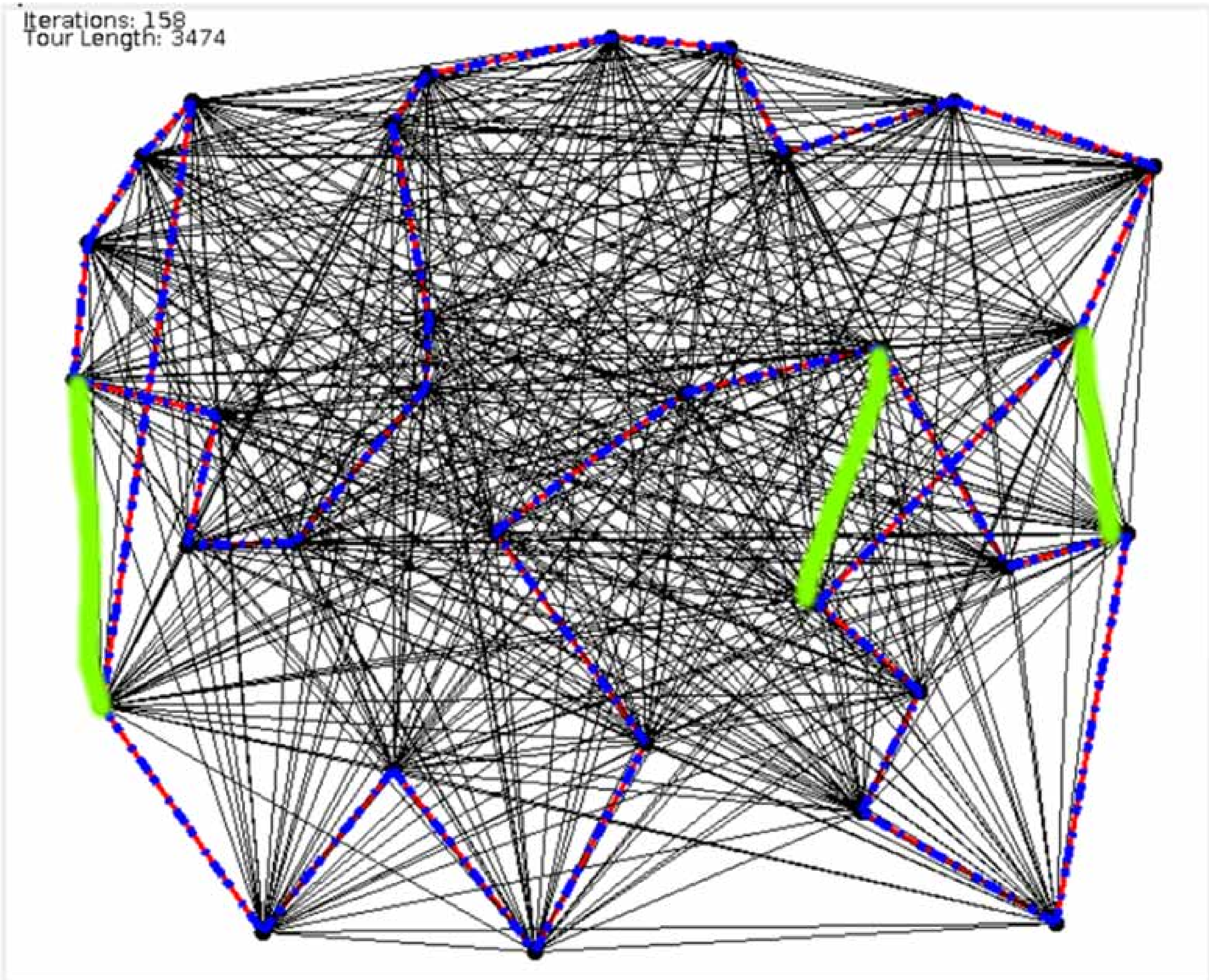
Human in the Loop



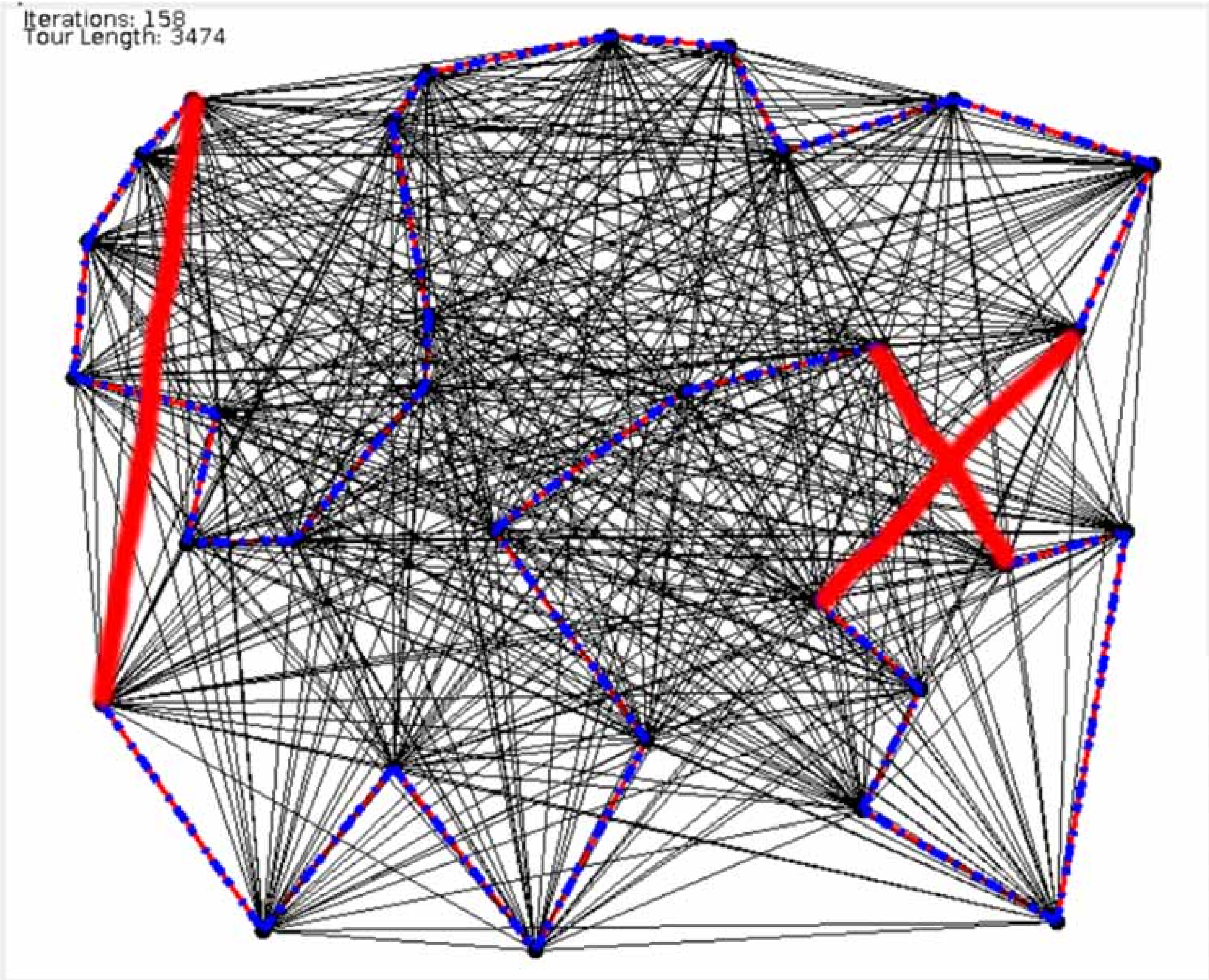
Bring in the Human



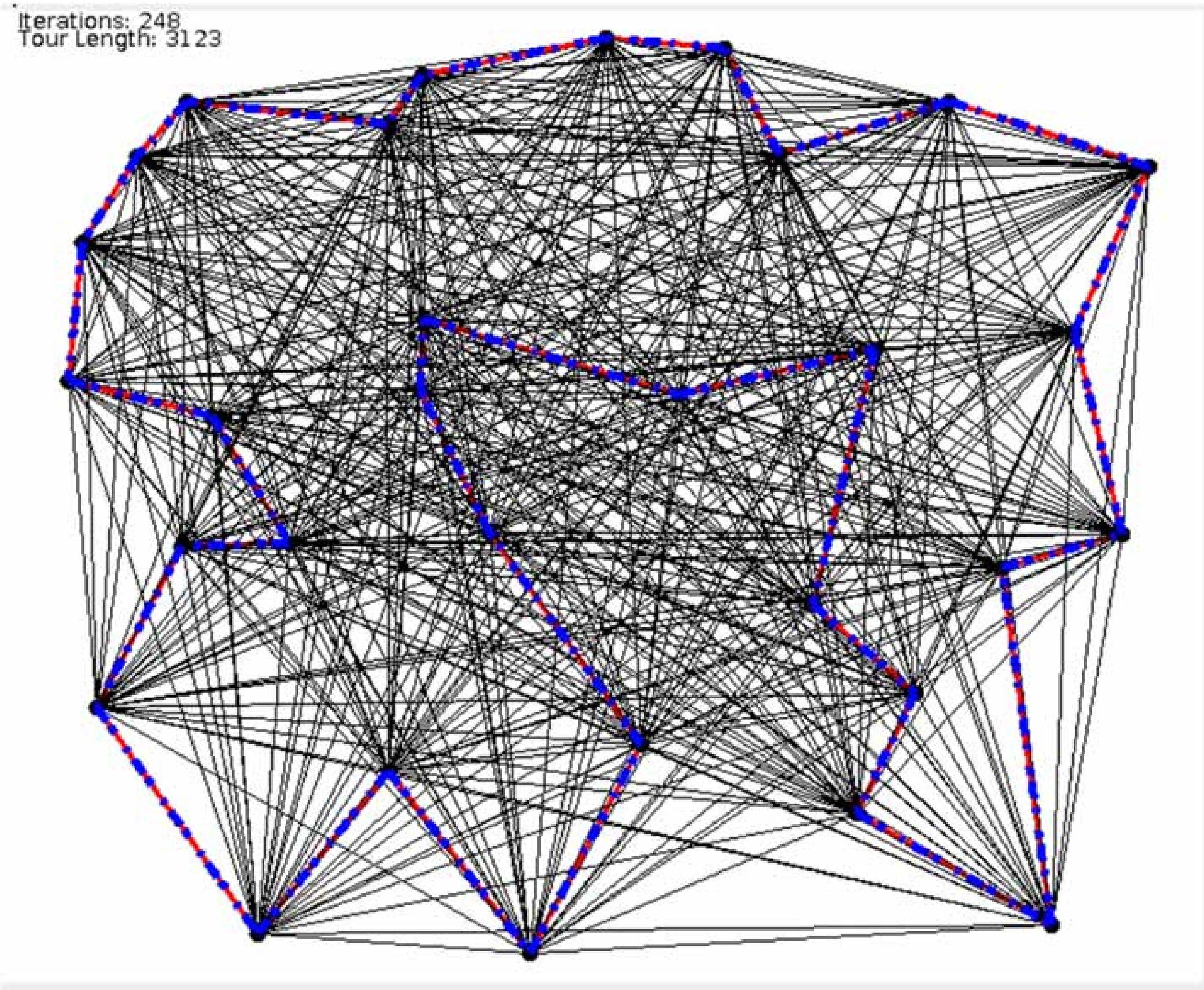
Add Pheromones



Remove Pheromones



Result:



Thank you!

Questions

Sample Questions (1)

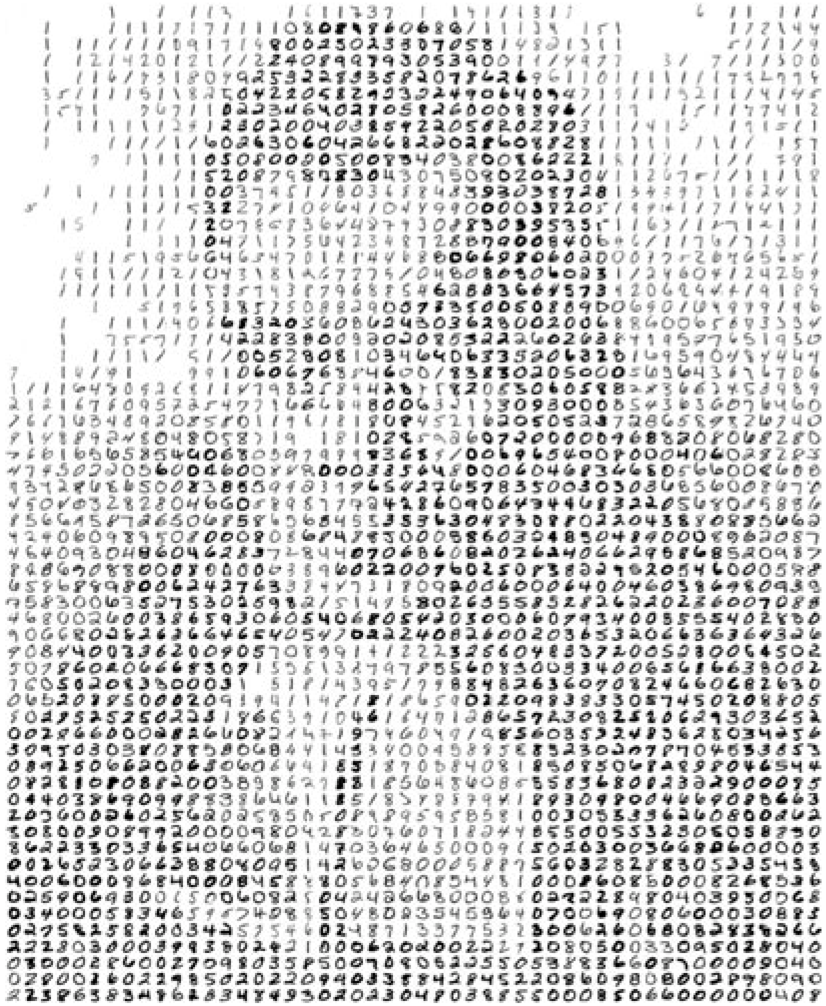
- Please explain the five mainstreams in ML!
- Why is it generally not easy to solve problems in health informatics?
- What is the model of a computational agent?
- Why is protein folding a hard problem?
- Explain why the study of human learning and machine learning can benefit from each other?
- What is a Pheromon and how does it work?
- In which areas are humans better than computers?
- What is the human kernel experiment?
- Why is simulated annealing interesting?
- Explain the Ant Colony Algorithm via pseudo code!
- Why should we study natural computing?

Appendix

"The contagion spread rapidly and before its progress could be arrested, sixteen persons were affected of which two died. Of these sixteen, eight were under my care. On this occasion I used for the first time the affusion of cold water in the manner described by Dr. Wright. It was first tried in two cases ... [then] employed in five other cases. It was repeated daily, and of these seven patients, the whole recovered."

Currie (1798)
Medical Reports on, the Effects of Water, Cold and Warm, as a Remedy in Fevers and Febrile Diseases

Appendix

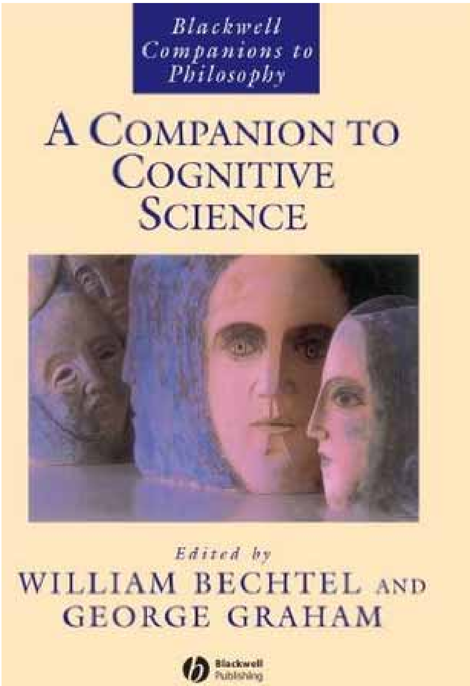


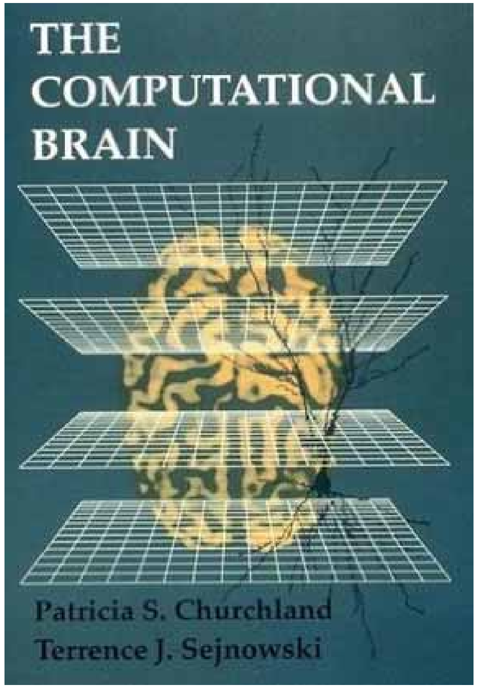
- Testing of novel Evolutionary algorithms:
  - Intelligent Water Drops
  - Bacteria Foraging Search
  - ...

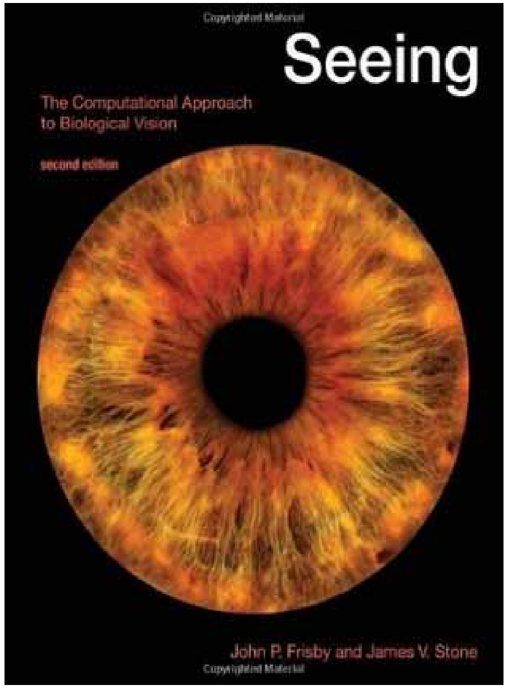
- EVOLKNO crowdsourcing platform to implement and test new algorithms:
  - Open Source data for Researchers to test algorithms
  - Evaluate quality, reusability and efficiency of algorithms

[16] Holzinger, K., Palade, V., Rabadan, R., & Holzinger, A. (2014). Darwin or lamarck? future challenges in evolutionary algorithms for knowledge discovery and data mining. In Interactive Knowledge Discovery and Data Mining in Biomedical Informatics (pp. 35-56). Springer Berlin Heidelberg.

Recommendable reading for further studies



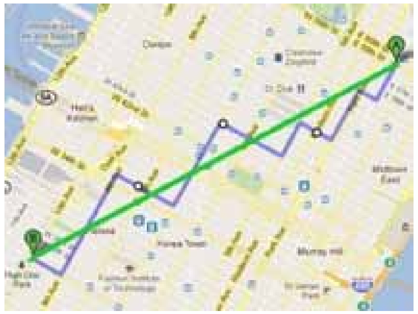




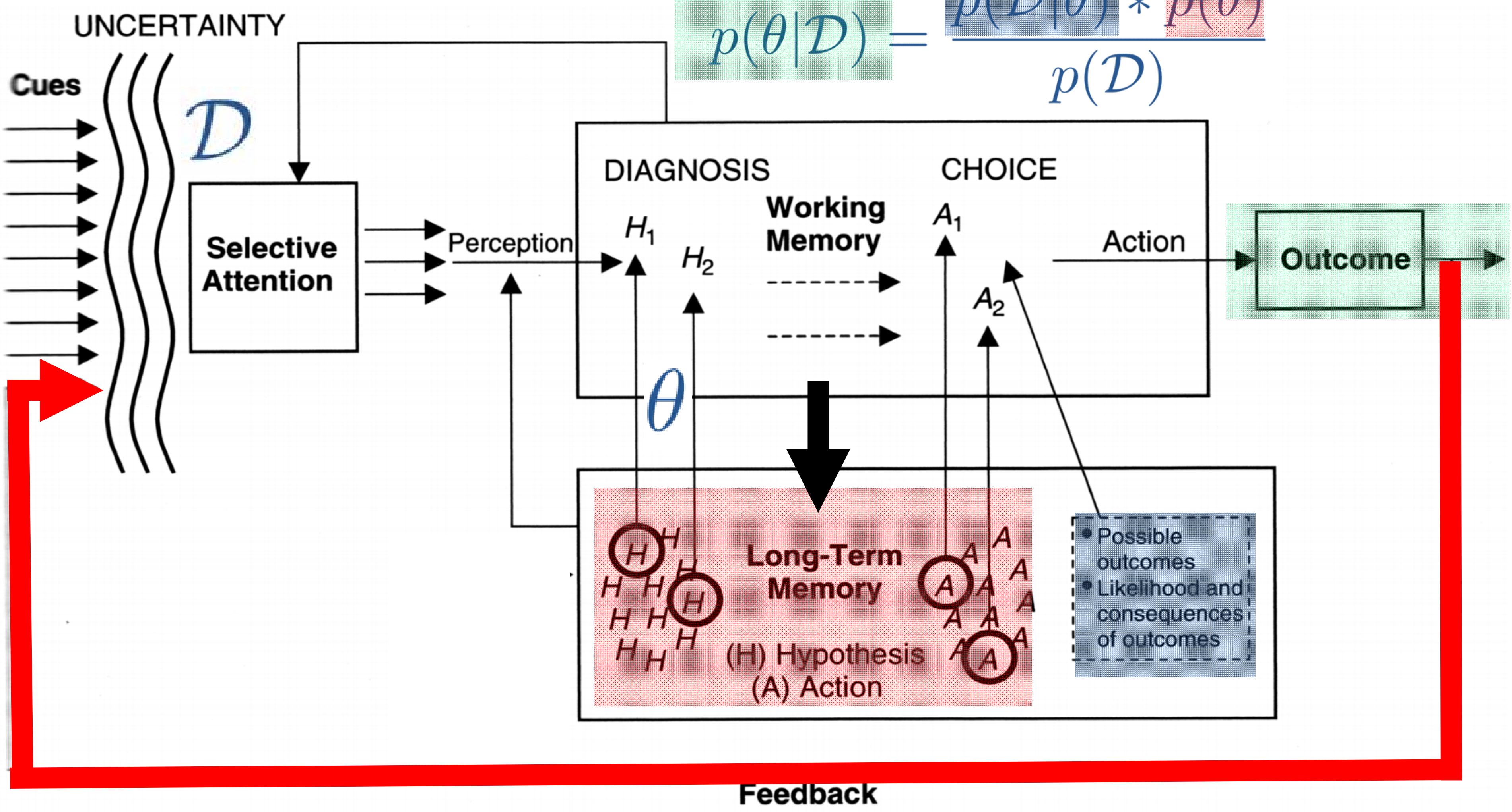
Answers to the Quiz Questions (1/2)

- 1 = This is a **chromosome –** in computation we call it a sequence of **information objects.** Each cell of any living creature has blueprints in the form of this chromosomes, which are strings of DNA and blocks of DNA, called ‘genes’, are responsible for the manifestation of traits, such as eye color, beard, etc.; Building blocks for chromosomes are proteins.
- 2 = This is a typical **naïve Bayes classifier:** An example E is classified to the class with the maximum posterior probability; wnb = weighted naïve Bayes, V denotes the classification given by the wnb, and is the weight of the attribute; The naive Bayes classifier combines this model with a decision rule. One common rule is to pick the hypothesis that is most probable; this is known as the maximum a posteriori or MAP decision rule.
- 3= This is the famous finding of Charles **Darwin: tree** of life. Darwin used the tree-structure in the context of his theory of evolution: Populations of individuals compete for limited resources; a fitness function is associated with each individual, which quantifies ability to survive; Parent populations reproduce to form offspring populations; and the traits of offspring are a combination of the traits of parents.

- 4=This is the experiment by Mnih et al (2015) "Google Deepmind": Human-level control through deep reinforcement learning, before the GO hype. They applied a deep network for playing an **Atari-Game.**
- 5=The **classification** experiment by Josh Tenenbaum, where he asks the question: How does the human mind get so much from so little?
- 6=Amazingly fascinating **big numbers:** We have 10^{80} elementary particles in the universe, multiplied by 10^{40} time steps since the big bang, we have 10^{120} possible computations in the universe – an amazing large number – BUT (big but!): one DNA molecule carries genetic information of the DNA with $3*10^9$ base pairs having 4^{3*10^9} combinations – which is a far larger number !!
- 7= **Distance measures**, Euclidean, Manhattan, Maximum; very important for similarity measures of vectors. The Manhattan distance is the simple sum of the horizontal and vertical components, whereas the diagonal distance might be computed by applying the Pythagorean theorem.



$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta) * p(\theta)}{p(\mathcal{D})}$$



Wickens, C. D. (1984) *Engineering psychology and human performance. Columbus (OH), Charles Merrill, Altered by Holzinger, A. (2017)*