The University of Birmingham

# Computational Intelligence for Biological Problems

Shan He

The Wellcome Trust VIP fellow

**School of Computer Science & Centre for Systems Biology**

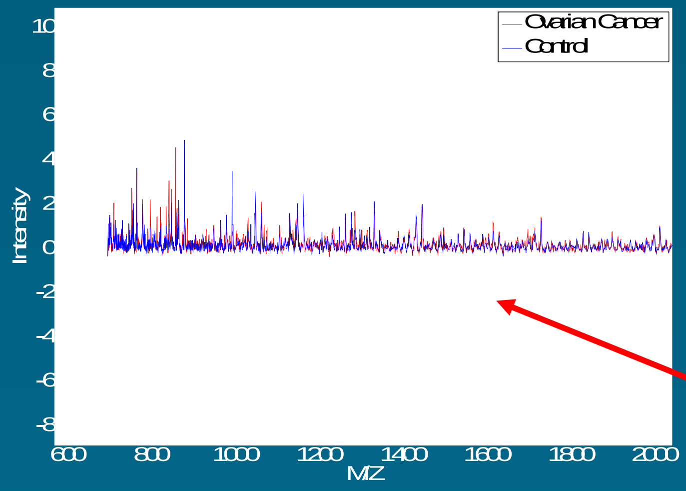
# Overview

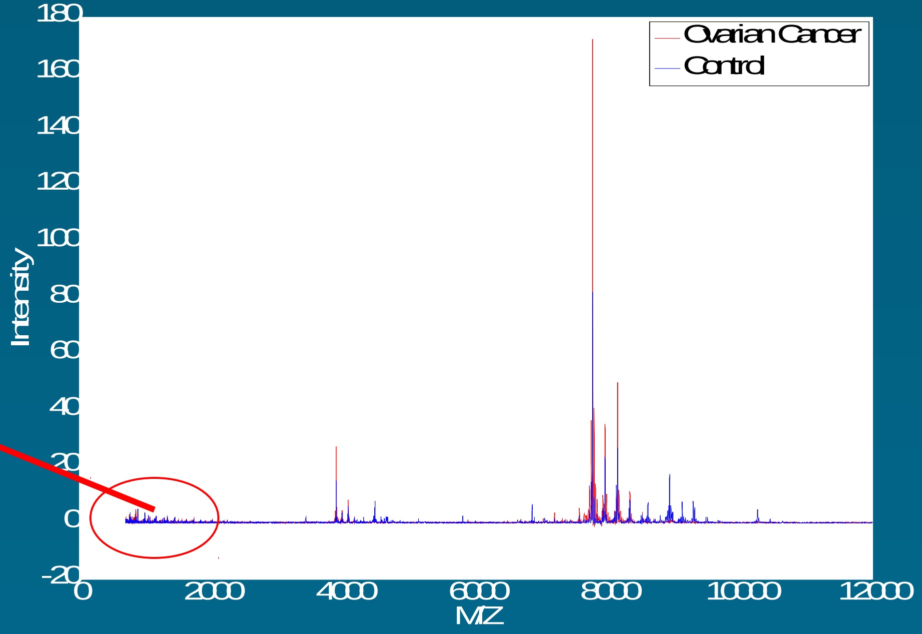
- Data mining for biomarker discovery from MS proteomics data
- Optimisation algorithm for the deconvolution of metabolomics data
- Agent-based modelling of the evolution of self-organising systems

# Profiling proteomics data

- Discover cancer biomarkers from mass spectrometry data
- High dimensionality
- Substantial noise
- Relatively small number of samples

# SELDI spectrum

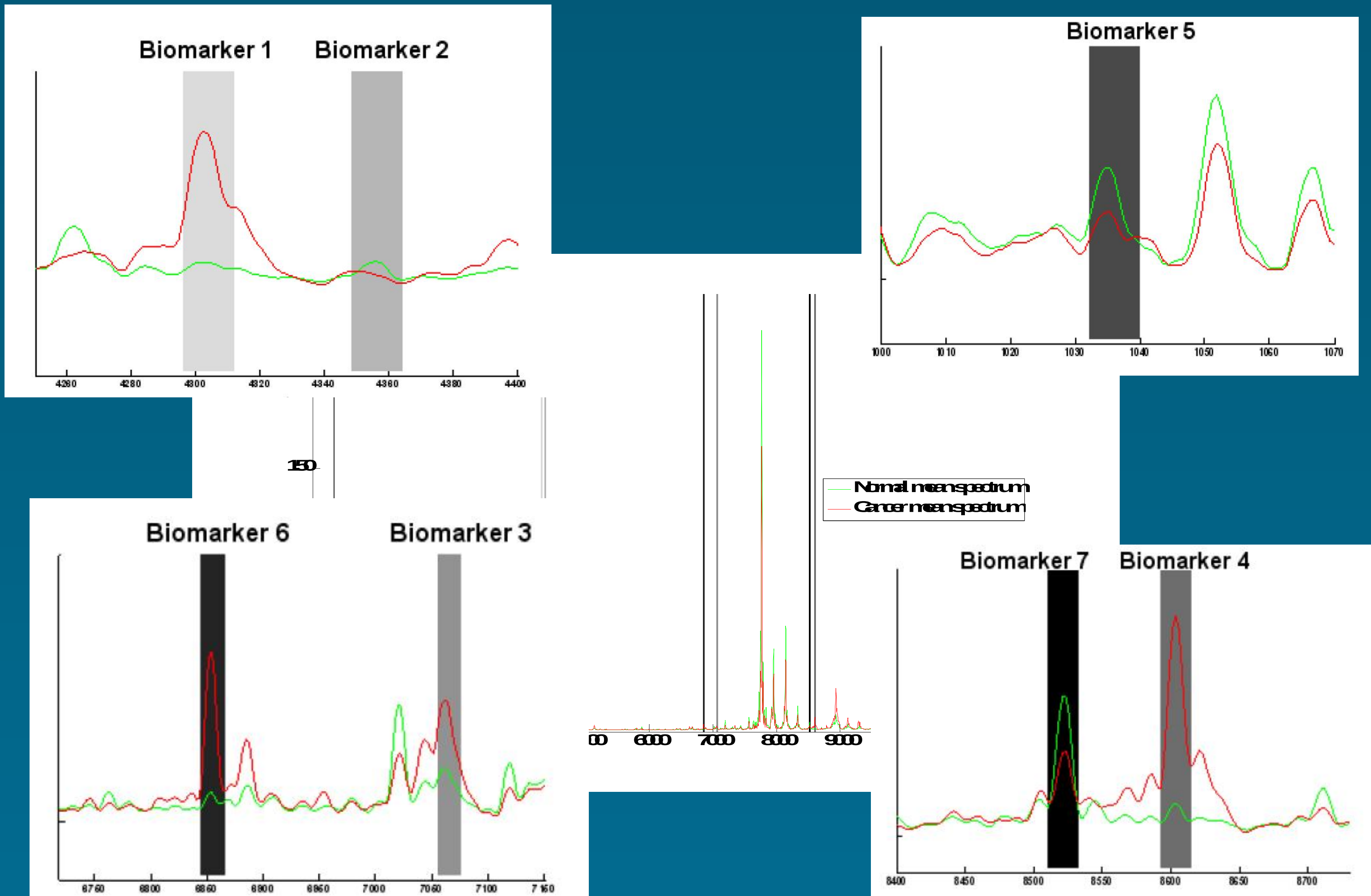


# My research in proteomics

- A novel proteomics data preprocessing algorithm
- A novel feature selection algorithm based on GSO
- A novel machine learning algorithm
- Ovarian, pancreatic and liver cancers

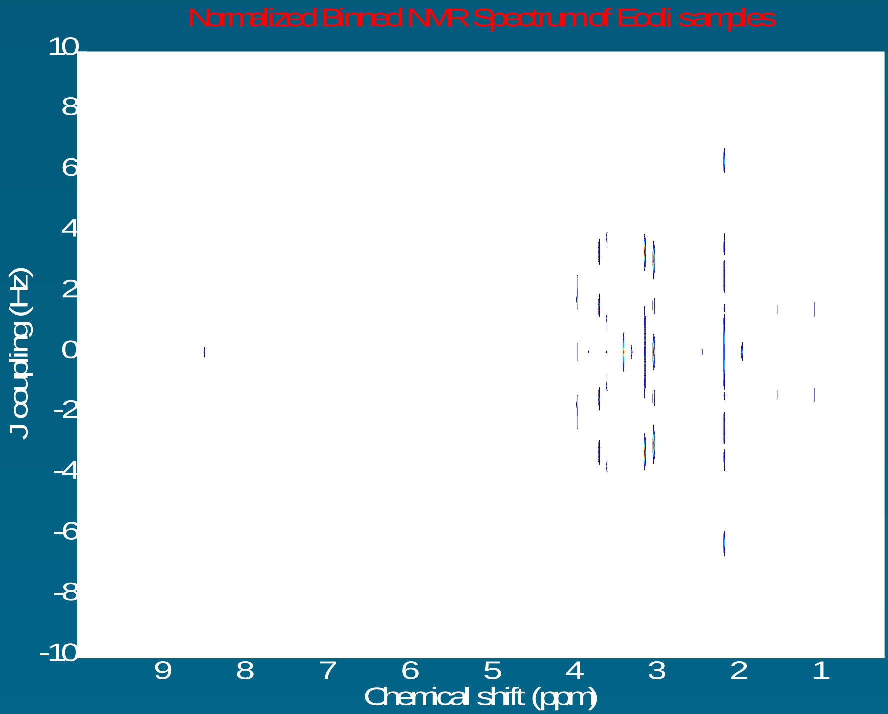
# Biomarkers we found

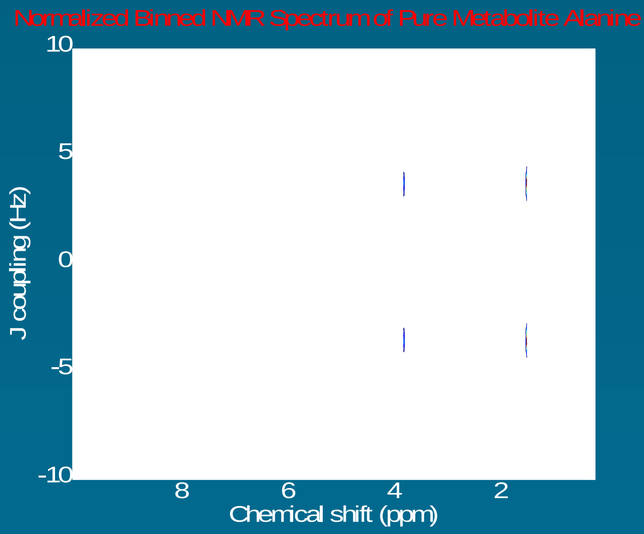


# Ident/Quant pure metabolites

- In collaboration with Dr. Mark Viant in School of Biosciences
- Objective: To indentify and quantify metabolites in biological samples based upon 2D J-resolved Nuclear Magnetic Resonance (NMR) spectra in conjunction with a metabolite library

# 2D J-res NMR spectra



# Our method

- Screen candidate pure metabolites
- Non-negative Constrained Linear Least-Square (NCLLS) algorithm
- The first identification/quantitication algorithm for 2D JRES spectra

# Results

- 3 simulated biological mixtures: cancer cell, mussel and Medaka fish embryo
- Screening: 100% sensitivity and 94.5% specificity
- Quantification: 5-6% error

# Self-organising aggregation

Pictures from NY Times

# Open problems

- Q1: How animals aggregate and coordinate?
- Q2: Why solitary animals evolved aggregation?
- Q3: How aggregation evolved through natural selection?

D. J. T. Sumpter. "The principles of collective animal behaviour". ***Philos Trans R Soc Lond B Biol Sci.*,** 361:5–22, 2006.

J. K. Parrish et al., "Complexity, pattern, and evolutionary trade-offs in animal aggregation". ***Science***, 284:99–101, 1999.

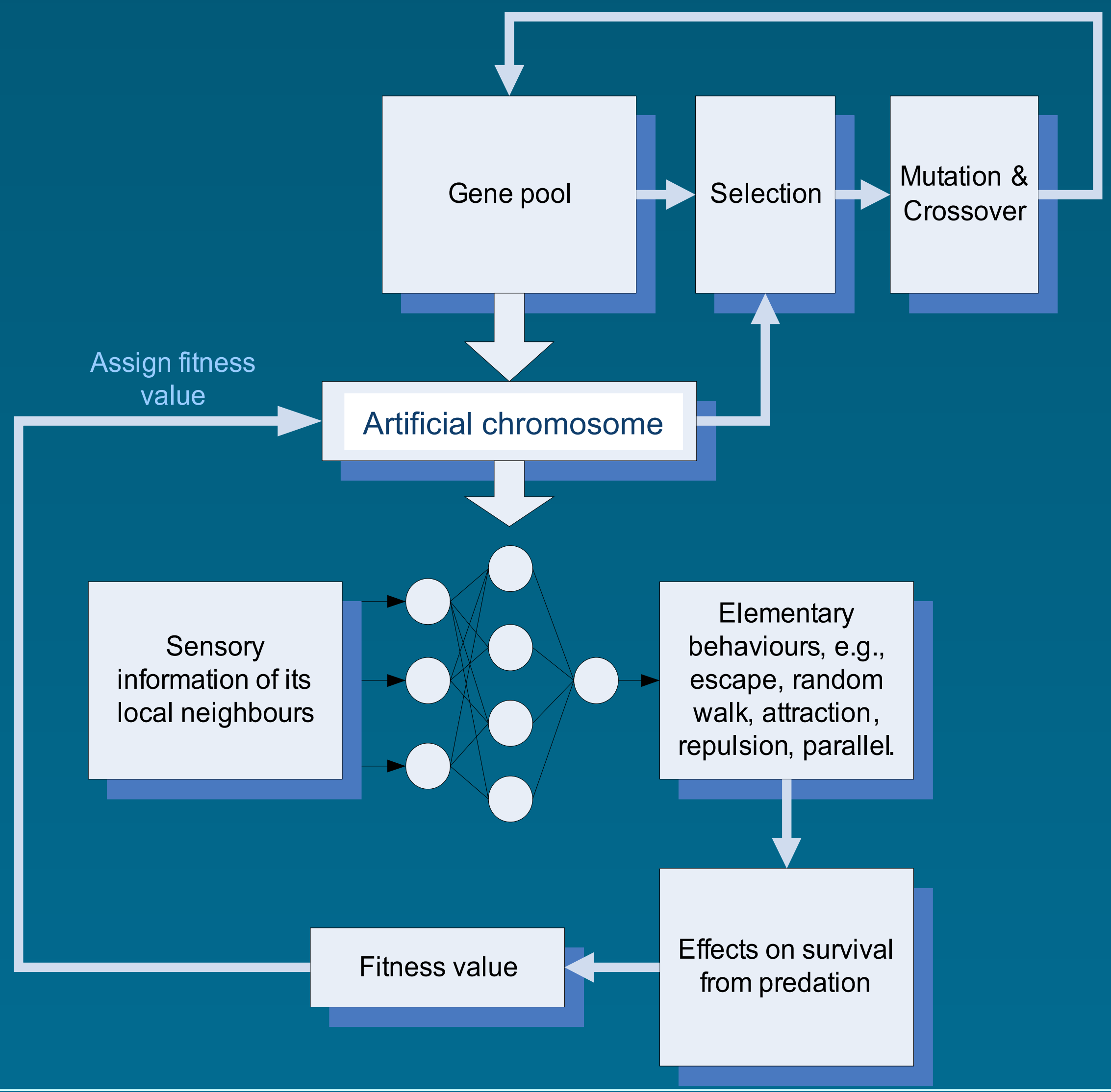
# Previous research

- Agent-based models – rule-based
- Possible mechanisms: Self-organisation
- A set of simple behavioural rules interacting locally
- Rules biologically plausible??
- Difficult to incorporate evolution dynamics

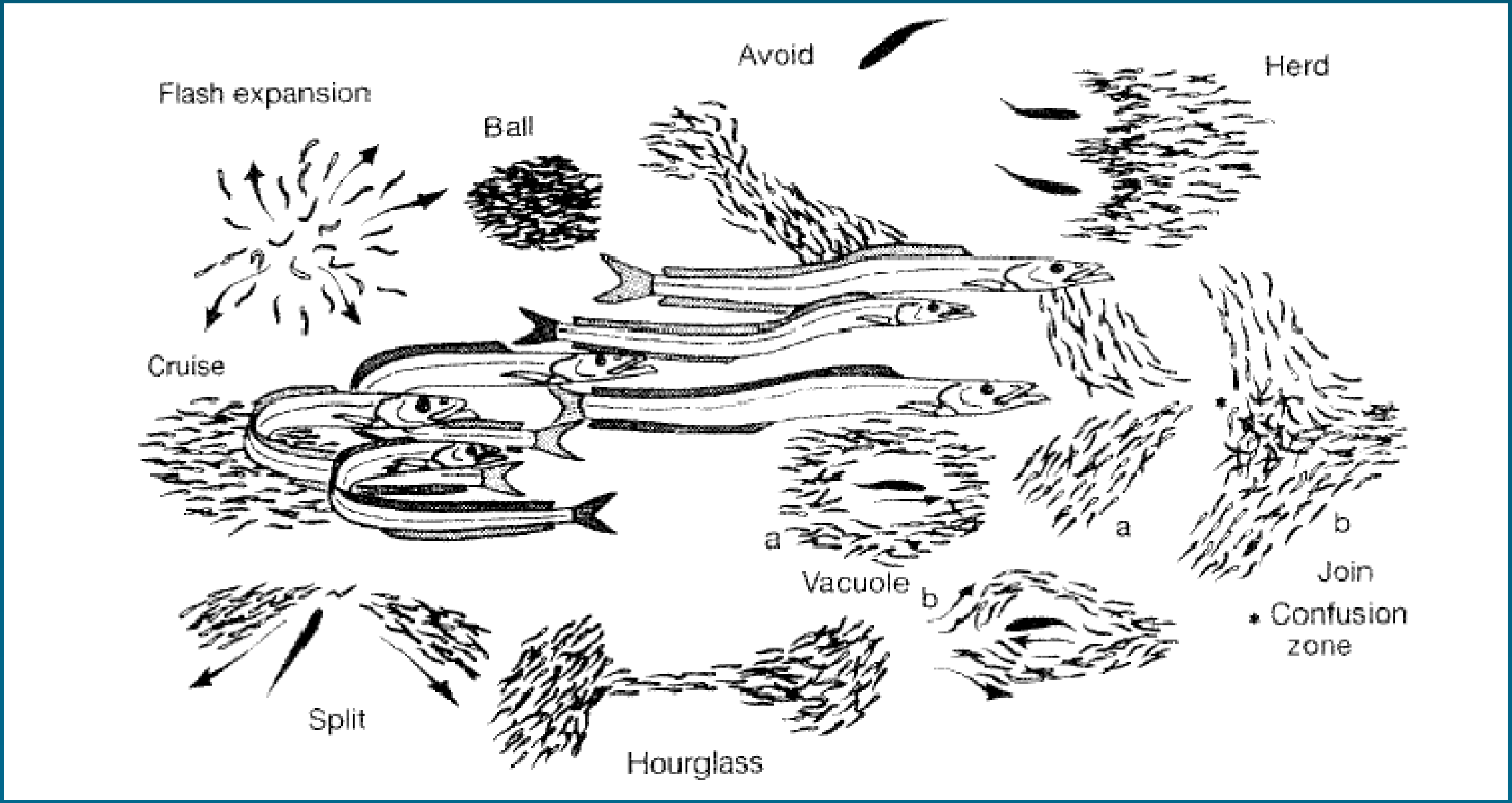
# My research

- Anti-predation → Animal aggregation
- Predator-prey games
- Agent-based modelling + Evolutionary Artificial Neural Networks (EANN)
- Each EANN controls an agent
- Local information → behaviours
- Level of selection

# Evolutionary process



# Aggregation patterns in nature




J. K. Parrish, S. V. Viscido, D. Grunbaum. “Self-Organized Fish Schools: An Examination of Emergent Properties”. ***Biol Bull***, Vol. 202, No. 3. 2002

# Main results

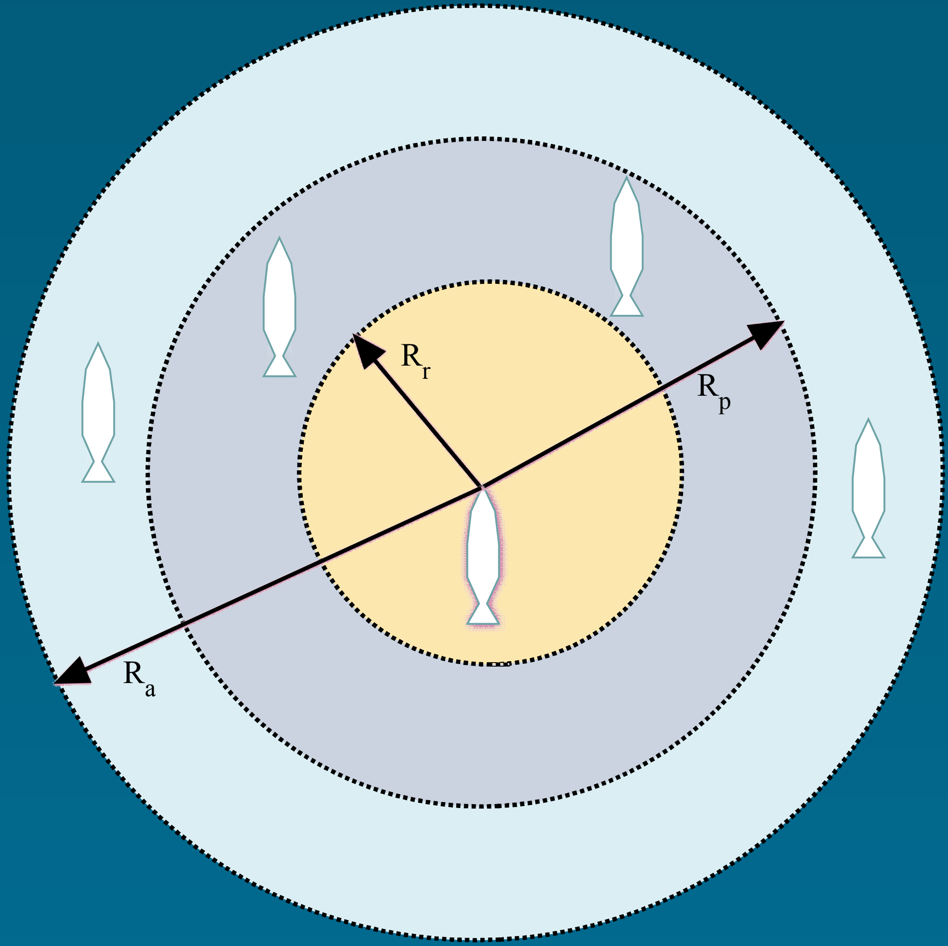
- Anti-predation is sufficient to evolve aggregation from solitary individuals by selection at individual level
- Open questions & major transitions
- Novel computational framework

# Surprising result!

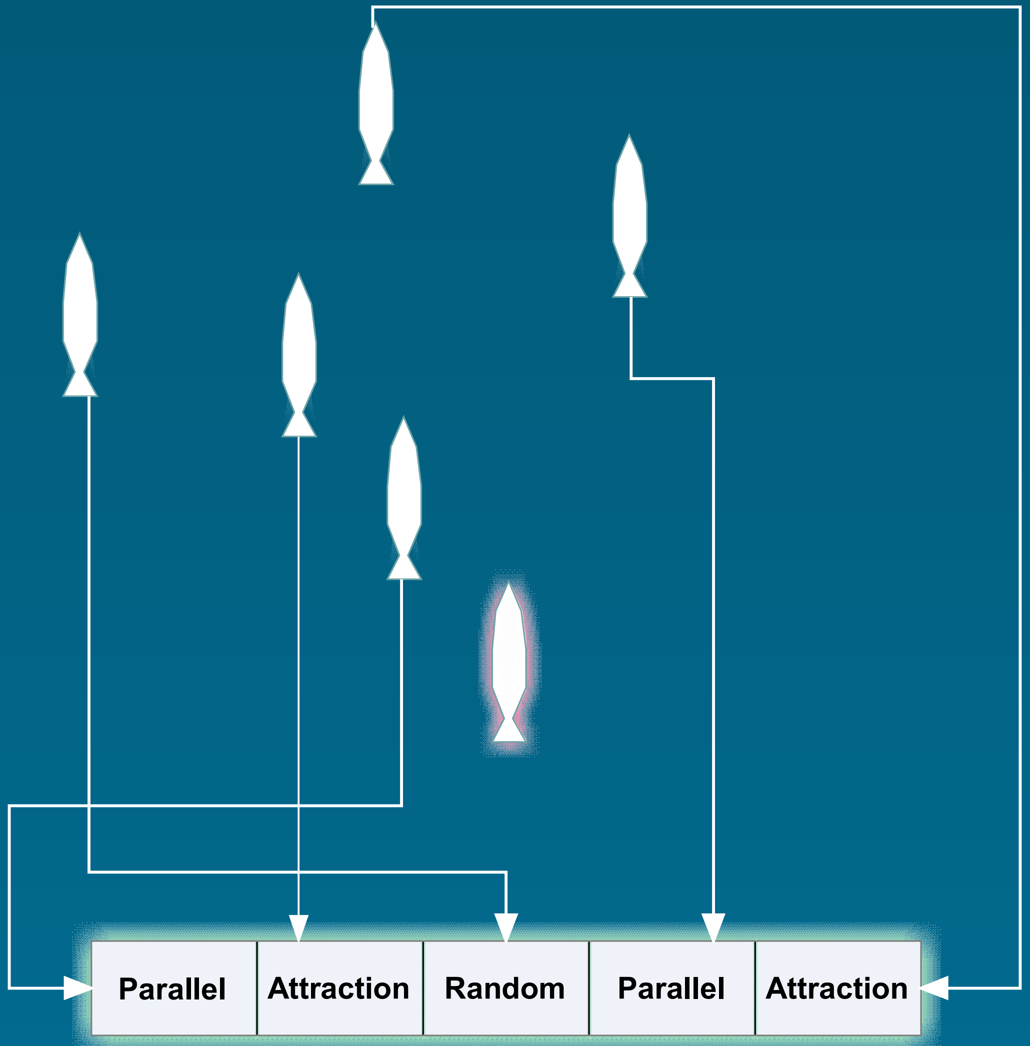
- Rule-extraction from EANNs
- Topological not metric interactions!!
- Topological interactions among flock of starlings

M. Ballerini, et al. “Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study”. ***Proceedings of the National Academy of Sciences***, 105, 2008.

# Metric interaction



# Topological interaction



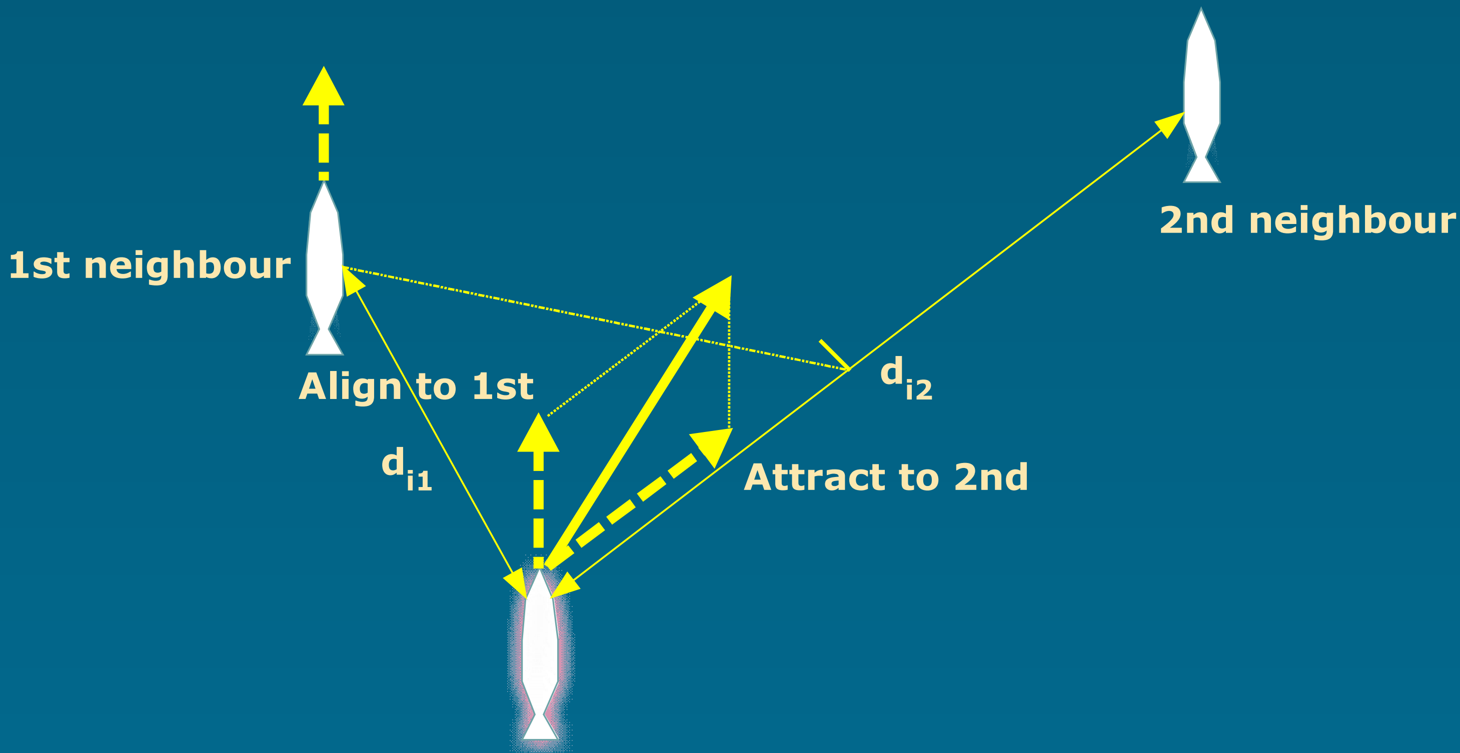

**ANN outputs**

# More surprising result!!

- Novel **non-metric-nor-topological** interaction  - never reported before!!!
- Interacting with only 2 neighbours

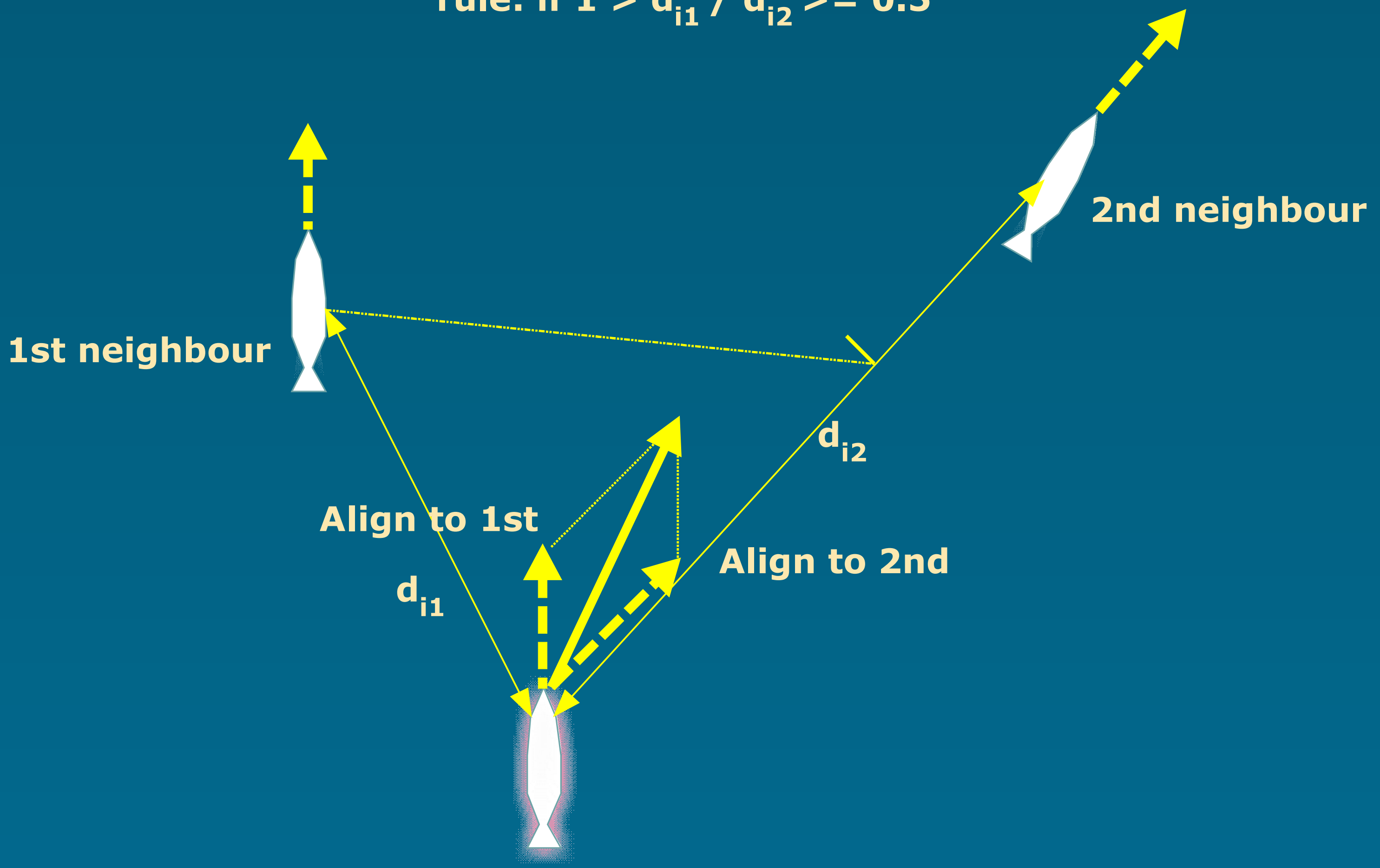
# Novel interaction rule

**rule: if $d_{i1} / d_{i2} < 0.5$**

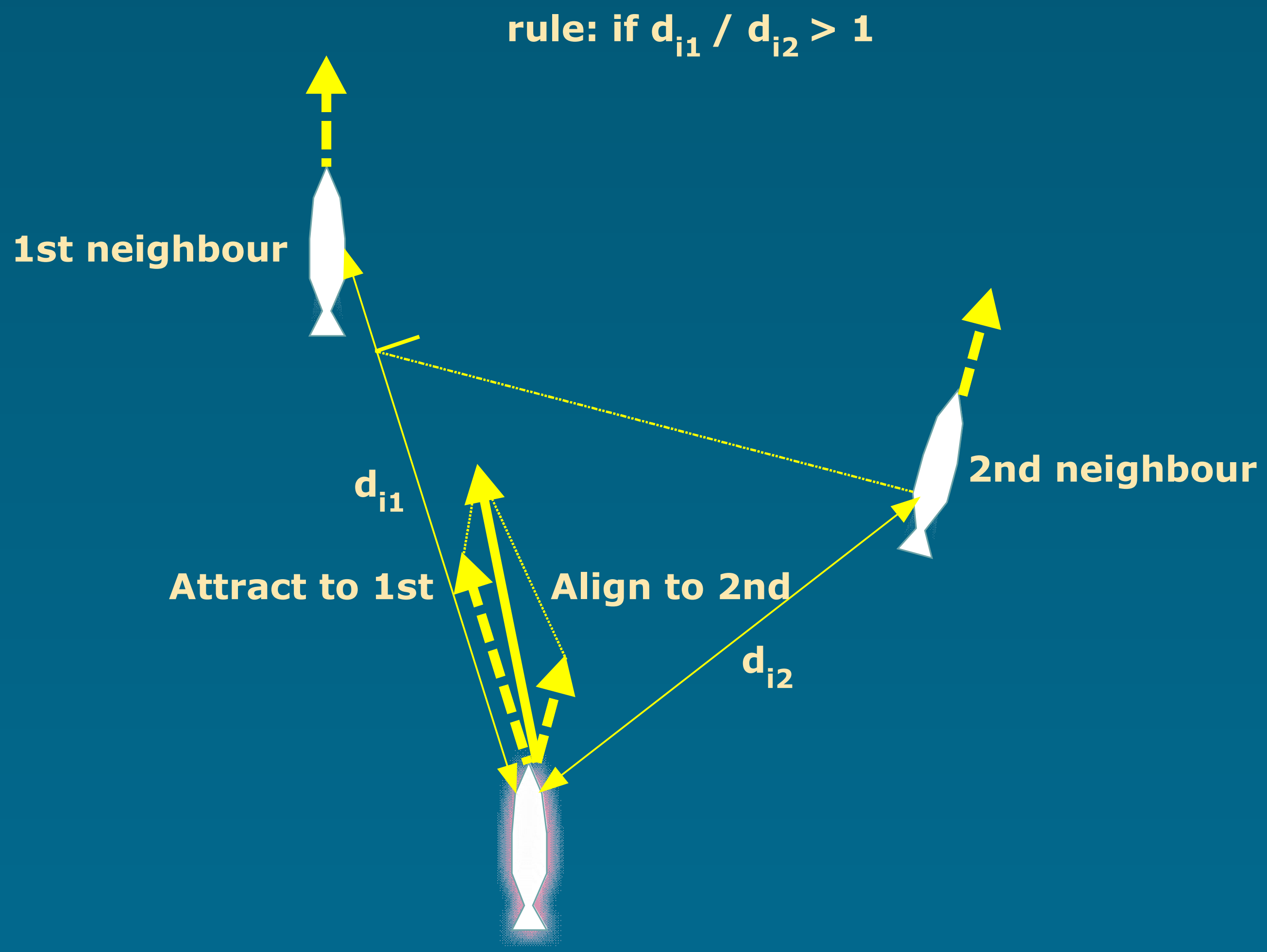


# Novel interaction rule

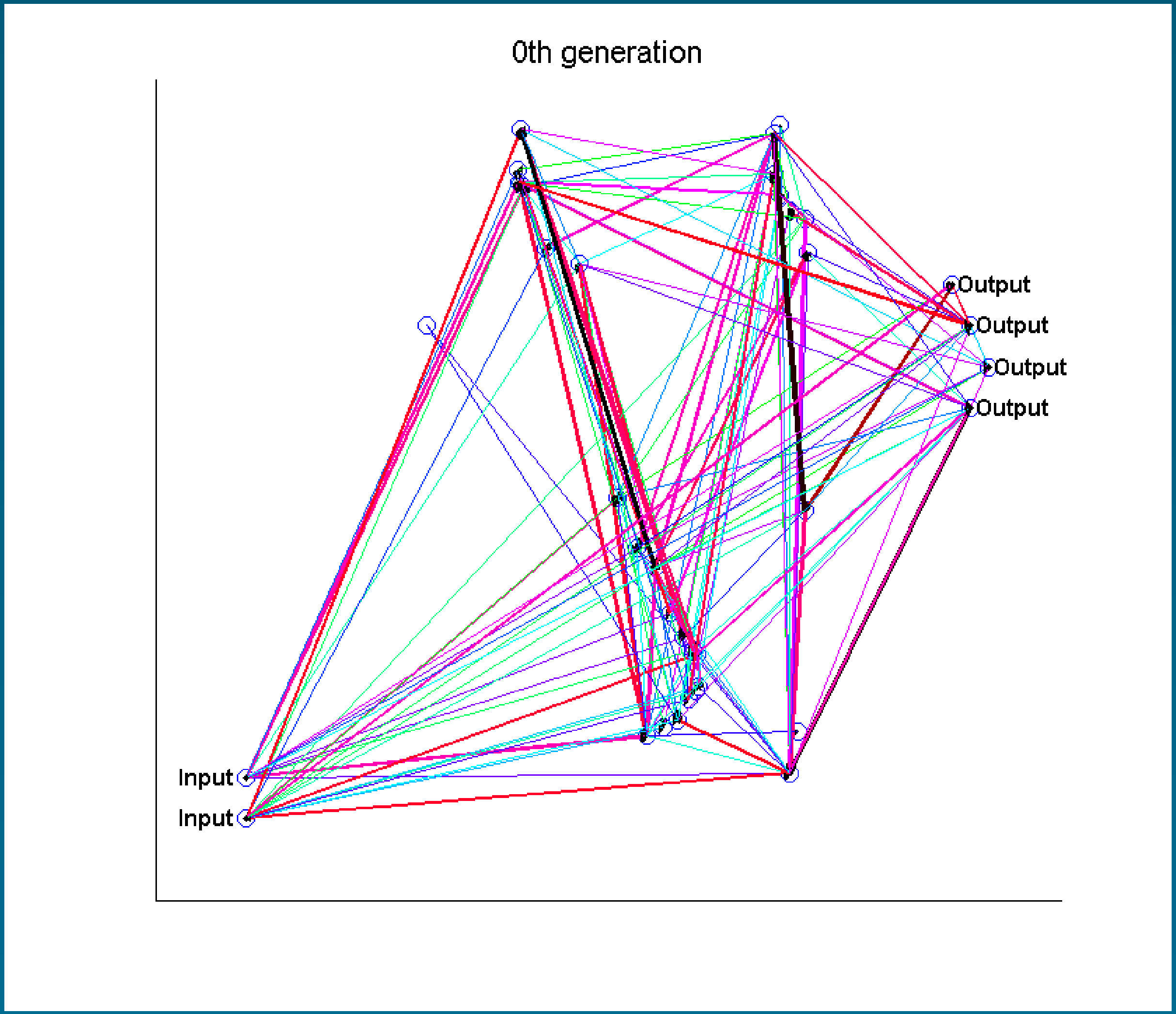
rule: if $1 > d_{i1} / d_{i2} >= 0.5$



# Novel interaction rule



# Visualisation of EANN



Thank you!

# Encoding scheme

- Marker-based encoding

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ... | start | key | initial_value | $k_1$ | $w_1$ | $k_2$ | $w_2$ | ... | $k_n$ | $w_n$ | end | ... |

where

| | |
|---|---|
| start | - Start marker. |
| key | - Node label used in the connections. |
| initial_value | - Output value of this node prior to first evaluation. |
| $k_i$ | - Connection source. |
| $w_i$ | - Connection weight. |
| end | - End marker. |