Bio-ICT Convergence: Filling the Gap Between Computer Science and Biology

Sara Montagna

sara.montagna@unibo.it

Ingegneria Due
Alma Mater Studiorum—Università di Bologna a Cesena

Academic Year 2010/2011

Outline

1. Context
   - BioICT-Convergence
2. From Bio to ICT
   - the chemical-inspired model of SAPERE
   - a crowd evacuation application
3. From ICT to Bio
   - a multilevel modelling framework – MS-BioNet
   - the use of metaheuristics for the parameter optimisation
   - evaluation on the analysis of *Drosophila Melanogaster* regionalisation
4. Conclusion and future works

Outline

BioICT-Convergence

From Bio to ICT

Designing and developing engineered system adopting the biological phenomena as a source of inspiration

From ICT to Bio

Using models, techniques and tools devised in computer science for addressing biological questions

Outline

Ecological Properties

Found in physics, chemistry, biology, human society . . .

Self-organisation

It is the process where a structure or pattern appears in a system without a central authority or external element imposing it through planning. This globally coherent pattern appears from the local interaction of the elements that make up the system.
The organisation is achieved in a way that is parallel (all the elements act at the same time) and distributed (no element is a coordinator).

Self-adaptation

Something, such as a device or mechanism, that changes so as to become suitable to a new or special application or situation

Self-optimisation, context-awarness, openness . . .

Self-organising patterns

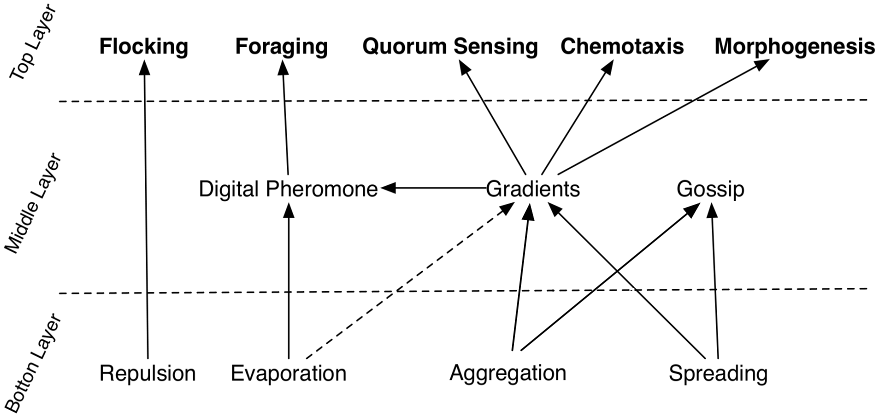
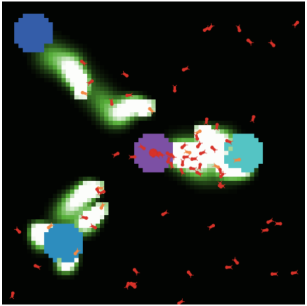


Figure: Patterns and their Relationships

Top Layer Patterns I

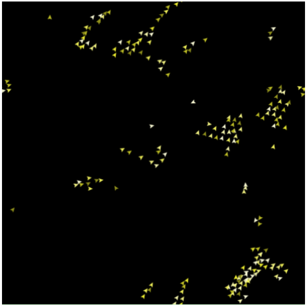
Foraging The activity where a set of ants collaborate to find the closest food to the nest.



Ant colonies use stigmergy communication, i.e. ants modify the environment through depositing a chemical substance called *pheromone*. This pheromone drives the behaviour of other ants in the colony.

Top Layer Patterns II

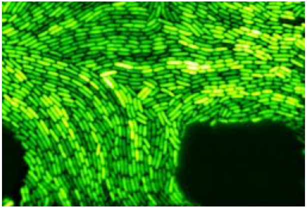
Flocking Behaviour of an herd of animals of similar size and body orientation.



Animals move en masse or migrate in the same direction and with a common group objective

Top Layer Patterns III

Quorum Sensing Type of intercellular signal used by bacteria to monitor cell density to coordinate certain behaviours (e.g. bioluminescence).



The bacteria constantly produce and secrete certain signaling molecules called auto-inducers. In presence of a high number of bacteria, the level of auto-inducers increases exponentially.

Top Layer Pattern IV: the Morphogenesis of Living Systems

Animal developmental steps

1. Fertilisation of one egg
2. Mitotic division
3. Cellular differentiation
    - diverse gene expression
4. Morphogenesis
    - control of the organised spatial distribution of the cell diversity

Each region of the developing organism expresses a given set of genes

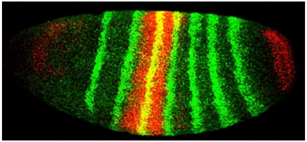


Figure: *Drosophila M.* segments

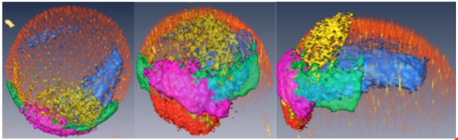
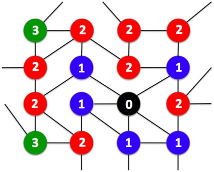


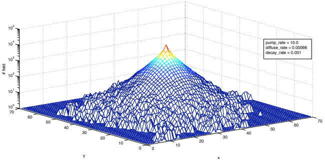
Figure: *Zebrafish* regionalisation

A Catalogue of Patterns

Middle Layer Patterns

Gradient The information is propagated in such a way that it provides an additional information about the sender's distance. In real systems gradients support long-range communication among biological entities (cells, bacteries, etc..) through local interaction.





Digital Pheoromone A digital pheromone is a mark, that is spread over the environment. Then, other ants beyond the communication range can receive the information generated by digital pheromones. Pheromones quickly evaporate.

A Catalogue of Patterns

Bottom Layer Patterns

- **Evaporation** Information evaporation or degradation.
- **Aggregation** Information fusion to produce a more useful information.
- **Spreading** Diffusion and dissemination of information over the environment or the direct communication among entities such as cells in a biological system.

The information in a real system might be a molecule, a data ...

Self-organising patterns

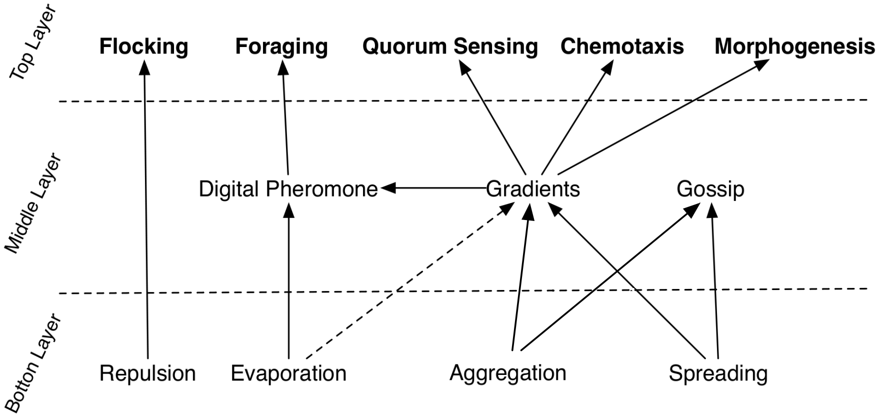


Figure: Patterns and their Relationships

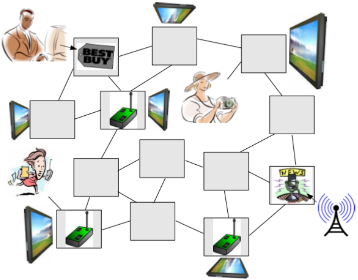
From Bio-* to Pervasive Computing Scenarios

Using the mechanisms that originate the spatial self-organising properties of real phenomena in the synthesis of pervasive systems

Pervasive Computing

Set of connected devices able to communicate

- large-scale distribution
- opennes
- context-awareness

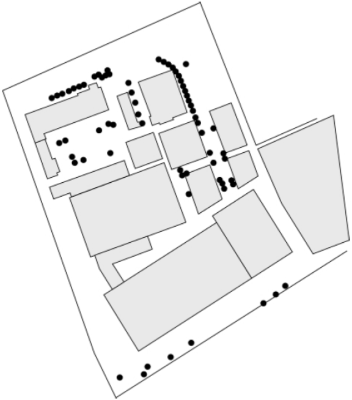


→ self-organisation and self-adaptation

Possible Real-World Scenarios of Pervasive Computing

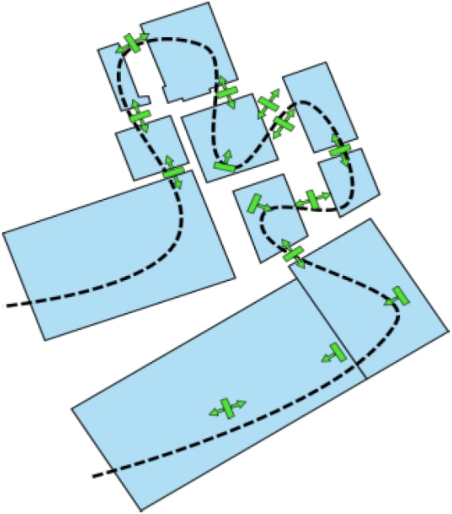
The crowd steering application scenario

A large scaled event area potentially consisting of multiple buildings which is populated by pervasive public displays



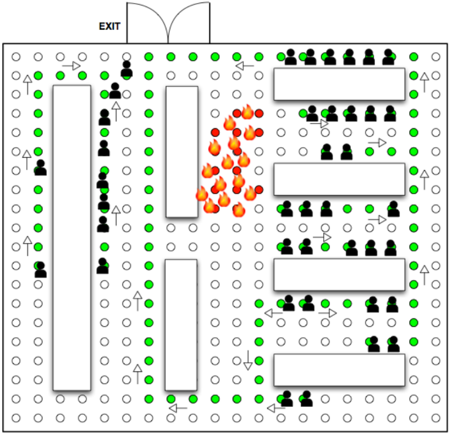
The Crowd Steering Application Scenario

Single user steering via public displays



The Crowd Steering Application Scenario

Crowd evacuation



Relevance for Pervasive Systems

Mapping . . . [15, 16]

- . . . individuals (services, requests) into biochemical species
- . . . the space into the network of compartments

Ecological properties useful in pervasive scenarios to coordinate users and services

Self-organisation

Users moving in the physical environment with PDAs/smart phones are reached by diffusing services

Self-adaptation

The best service is selected over time

Self-optimisation

Unused services get disposed

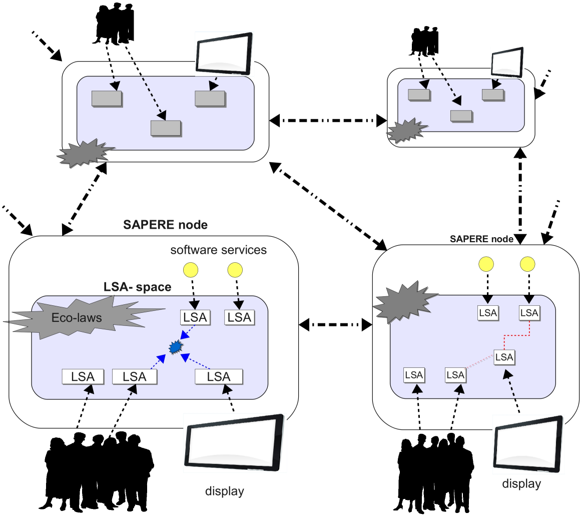
Openness

The system can deal with incoming new services and requests

Context-awareness

Local field value in a node depends on the state of the surrounding nodes

The SAPERE model and architecture



The SAPERE Model and Architecture: External Agents

All the components forming the ecosystem are modelled as agents

- Humans perceiving/acting over the system directly or through PDAs
- Pervasive devices (displays, sensors)
- Software services

Agents reify their state to the system through ...

... *Live Semantic Annotations* (LSA)

- Semantic representation of the agent's relevant interface/behavioural/configuration
- Reflect the current situation and context of the component they describe

The SAPERE Model and Architecture: the Infrastructure

As soon as a component enters the ecosystem, its LSA will be automatically created and injected in ...

... the SAPERE substrate

- Shared space where all LSAs live and interact
- Topology
  - structured as a network of LSA-spaces, each hosted by a node of the SAPERE infrastructure
  - keeping the set of LSAs of the components around—proximity of two LSA-spaces implies direct communication abilities

The SAPERE Model and Architecture: the Eco-laws

Each LSA-space embeds the basic laws of the ecosystem

The *eco-laws*

- They rule the activities of the system by properly evolving the populations of LSAs
- They define sorts of virtual *chemical reactions* among the LSAs, enforcing coordination of data and services.

The SAPERE Model and Architecture: Eco-laws & LSAs

Data and services. . .

- Are represented by their associated LSAs
- Are sorts of chemical reagents in an ecology
- Their interactions and composition occur via chemical-like reactions, i.e., pattern-matching between LSAs

Such reactions can contribute to. . .

- Establish virtual chemical bonds between entities
- Produce new components
- Diffuse LSAs as in biochemical systems

Coordination, Self-organisation and Adaptivity in the SAPERE Framework

- They are not bound inside the capability of individual components
- They rather emerge in the overall dynamics of the ecosystem
- They are ensured by the fact that any change in the system will reflect in the firing of some eco-law possibly leading to the creating/removal/modification of LSAs

The Role of Simulation for Designing Pervasive Systems

Models and simulation for supporting the design of pervasive systems

Possibility to. . .

- Experiment the idea of exploiting bio-inspired ecological mechanisms
- Showing through simulation the overall behaviour of a system designed on top of eco-laws
- Elaborate what-if scenarios

To capture the whole complexity of the SAPERE approach the model has to support the abstraction of . . .

- Highly dynamic environment composed of different, mobile, communicating nodes
- Autonomous agents

They might be programmable through a set of chemical rules

In Literature: the Agent-based Model

Agent-based model is a specific individual-based computational model for studying macro emergent phenomena through the definition of the system micro level which is modelled as a collection of interacting entities.

- MAS provides designers and developers with. . .
  - Agents
    ...a way of structuring a model around autonomous, heterogeneous, communicative, possibly adaptive, intelligent, mobile and. . . entities
  - Environment
    ...a way of modelling an environment characterised by a topology and complex internal dynamics
- MAS gives methods to. . .
  - model individual structures and behaviours of different entities
  - model local interactions among entities and entities-environment
  - model the environment structures and dynamics

On the Adoption of ABM

Advantages

1. Quite natural as soon as the pervasive system itself is engineered adopting the agent paradigm
2. There are several works which apply this approach in different contexts, from social systems [1] to biological systems [6]
3. The environment is also a first class abstraction whose structure, topology and dynamic can be explicitly modelled
4. Different simulation frameworks available: MASON [14], Repast [12], NetLogo [17] and Swarm [13]

Drawbacks

1. ABM does not normally provide a way to define the behavioural rules in terms of chemical laws

In Literature: Formal Models and Bio-Chemical Simulators

- Based on computational models
  - stochastic process-algebras
  - Petri-Nets
- Promote a view of "molecules as concurrent processes"
- Simulated on top of SPIM (stochastic π-calculus), BlenX, Bio-PEPA [2] and BetaWB [3]
- Ground on Gillespie's characterisation of chemistry as CTMC
  - concentration is evolved "exactly" as in chemistry

Gillespie "direct" simulation algorithm [5]

1. Compute the markovian rate $r_1, \ldots, r_n$ of reactions, let R be the sum
2. Choose one of them probabilistically, and execute its transition
3. Proceed again with (1) after $\frac{1}{R} * ln\frac{1}{\tau}$ seconds, with $\tau = random(0,1)$

On the Adoption of Bio-Chemical Simulator

Advantages

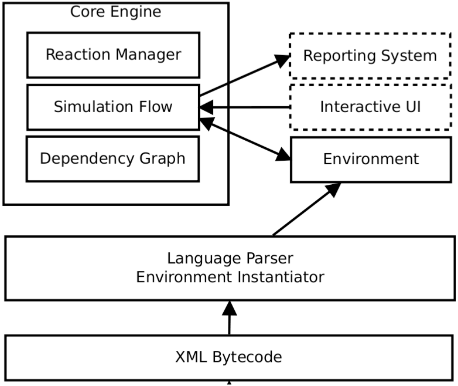
1. Helps for explicitly model the eco-laws

Drawbacks

1. Few simulators allow to define a multi-compartment topology
2. No one provides facilities to move compartments inside an external environment
3. All compartments are subject to the same set of laws, which are chemical reactions

Alchemist

- It faces natively the model requirements
- It takes the best of both approaches
- It implements an optimised version of the Gillespie's SSA, the Next Reaction Method [4]



The Gillespie's SSA: Premises

- *Well-stirred* system of molecules of N chemical species $\{S_1, ..., S_N\}$
- The molecules interact through M reactions $\{R_1, ..., R_M\}$
  - only *unimolcular* and *bimolecular* reactions are considered
  - *trimulecolar, reversible...* are modelled as a sequence of reactions
- Let $\mathbf{X}(t) = (X_1(t), ..., X_N(t))$ be the state of the system
  - $X_i(t)$ is the number of S_i molecules in the volume at time t
  - ν_j is the state change vector
- Let c_j be the reaction probability rate constant for R_j
- Let $a_j(\mathbf{x})dt$ be the *propensity function* for R_j as the probability, given $\mathbf{X}(t) = \mathbf{x}$, that an R_j will occur in $[t, t + dt)$

$$S_1 \xrightarrow{c_j} \text{products: } a_j(\mathbf{x})dt = c_j x_1$$

$$S_1 + S_2 \xrightarrow{c_j} \text{products: } a_j(\mathbf{x})dt = c_j x_1 x_2$$

$$2S_1 \xrightarrow{c_j} \text{products: } a_j(\mathbf{x})dt = c_j \frac{1}{2} x_1 (x_1 - 1)$$

The Gillespie's SSA: the *Direct Method*

1. Initialise the time $t = t_0$ and the system's state $\mathbf{x} = \mathbf{x}_0$
2. With the system in state $\mathbf{x}$ at time t, evaluate
   - all the $a_j(\mathbf{x})$
   - their sum $a_0(\mathbf{x}) \equiv \sum_{j'=1}^{M} a_{j'}(\mathbf{x})$
3. Draw two random numbers r_1 and r_2 from the uniform distribution in the unit interval and take

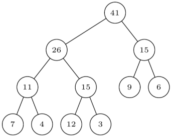
$$\tau = \frac{1}{a_0(\mathbf{x})} ln\left(\frac{1}{r_1}\right)$$

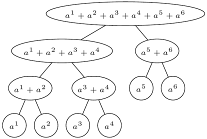
$$j = \text{the smallest integer satisfying} \sum_{j'=1}^{j} a_{j'}(\mathbf{x}) > r_2 a_0(\mathbf{x})$$

4. Effect the next reaction by replacing $t \leftarrow t + \tau$ and $\mathbf{x} \leftarrow \mathbf{x} + \nu_j$
5. Record $(\mathbf{x}, t)$ as desired
6. Return to step 2, or else stop

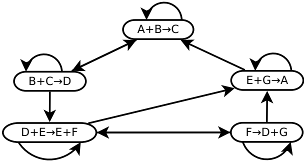
Efficient Exact Stochastic Simulation: the *Enhanced Direct Method* [4]

1. To select the next reaction it uses a binary tree





2. To update the propensity function it uses the dependency graph



Computational complexity of the Direct Method O(n)

Computational complexity of the Enhanced Direct Method O(log(n))

Outline

A Crowd Evacuation Application

The contest

- A museum composed of rooms, corridors, and exit doors
- The surface is covered by interconnected sensors detecting
  - a fire
  - the presence of people
- Each visitor has a PDA

When a fire breaks out, PDAs – which interact with sensors – must show the direction towards an exist, along a safe path.

PDA will...

Distance tend to guide people across the shortest path available to the nearest exit

Fire consider a path less safe if it passes near fire

Crowd consider a path less acceptable if it is overcrowded

The Model

- Biochemical species

  $\langle \mathtt{source}, \mathit{type}, \mathit{max}\rangle, \langle \mathtt{grad}, \mathit{type}, \mathit{value}, \mathit{max}\rangle, \langle \mathtt{info}, \mathit{type}, \mathit{value}, \mathit{timestamp}\rangle$

- Eco-laws for building the fire, exit and crowding gradients

$$\langle \mathtt{source}, T, M\rangle \xrightarrow{R_{init}} \langle \mathtt{source}, T, M\rangle, \langle \mathtt{grad}, T, 0, M\rangle$$
$$\langle \mathtt{grad}, T, V, M\rangle \xrightarrow{R_s} \langle \mathtt{grad}, T, V, M\rangle, +\langle \mathtt{grad}, T, \mathtt{min}(V{+}\#D, M), M\rangle$$

$$\langle \mathtt{grad}, T, V, M\rangle, \langle \mathtt{grad}, T, W, M\rangle \rightarrow \langle \mathtt{grad}, T, \mathtt{min}(V, W), M\rangle$$

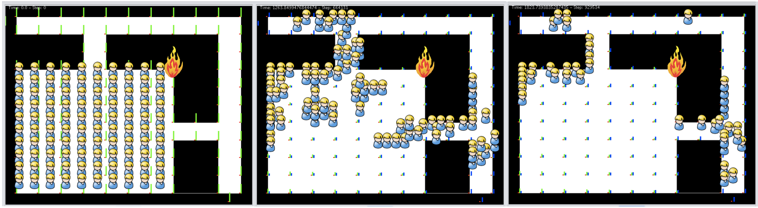
- Eco-laws for computing the attractiveness values

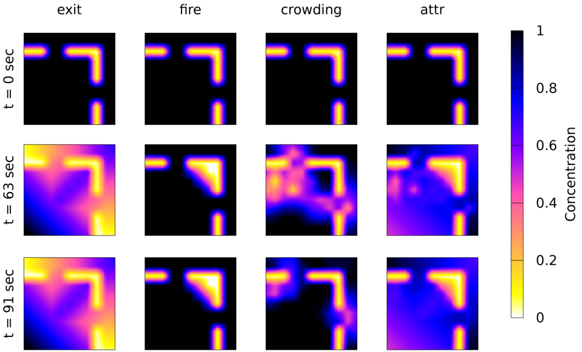
$$\langle \mathtt{grad}, \mathtt{exit}, E, Me\rangle, \langle \mathtt{grad}, \mathtt{fire}, F, Mf\rangle, \langle \mathtt{grad}, \mathtt{crowd}, CR, TS\rangle \xrightarrow{R_{att}}$$
$$\langle \mathtt{grad}, \mathtt{exit}, E, Me\rangle, \langle \mathtt{grad}, \mathtt{fire}, F, Mf\rangle, \langle \mathtt{info}, \mathtt{crowd}, CR, TS\rangle,$$
$$\langle \mathtt{info}, \mathtt{attr}, (Me - E)/(1 + (Mf - F) + (Mc - C), \#T\rangle$$

$$\langle \mathtt{info}, \mathtt{attr}, A, TS\rangle, \langle \mathtt{info}, \mathtt{attr}, A2, TS{+}T\rangle \rightarrow \langle \mathtt{info}, \mathtt{attr}, A2, TS{+}T\rangle$$

- People ascend the attractiveness gradient

Simulation Results





Outline

How to Model Morphogenesis?

Computational model requirements

1. Multi-compartment / multi-level model
    - for reproducing the interactions and integrations of the system components at cellular and intracellular level
2. Diffusion / Transfer
    - for studying the effects of short and long range signals
    - for modelling the compartment membrane
3. Stochasticity
    - for capturing the aleatory behaviour characteristic of those systems involving few entities

Ad-hoc Framework to Tackle Scenarios of Dev. Bio.

A new simulator based on computational models

- MS-BioNet
  - naturally supporting scenarios with many compartments
  - use state-of-the-art implem. techniques for the simulation engine
  - ground on Gillespie's characterisation of chemistry as CTMC

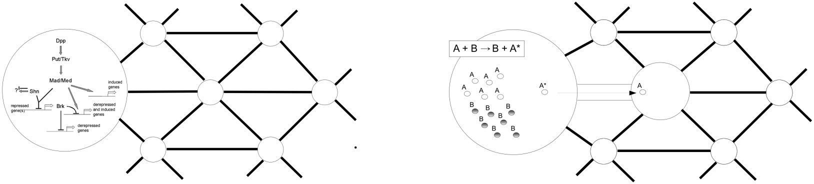
A module for parameter tuning

- Parameter tuning as an optimisation problem
  - searching the solution with metaheuristics

MS-BioNet

MS-BioNet's Conceptual levels [9]

1. Computational Model: graph of compartments, with transfer reactions



2. Surface Language: systems as logic-oriented description programs
   - system structure
   - inner chemical behaviours
3. Simulation Engine: implementation of Gillespie SSA [5]
   - reproducing the exact chemical evolution/diffusion of substances

Metaheuristics for Parameter Tuning in Comp. Bio.

Parameter tuning in Computational Biology

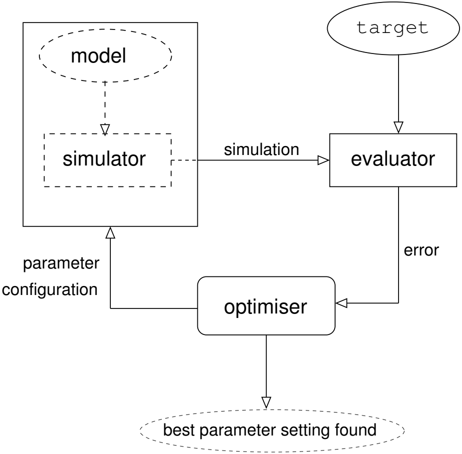
- Given the model structure and a set of target data
- Finding the values for model parameters so as to reproduce the system behaviour

The idea [7]

- Transforming parameter tuning into an optimisation problem solved with metaheuristics

Framework's Architecture

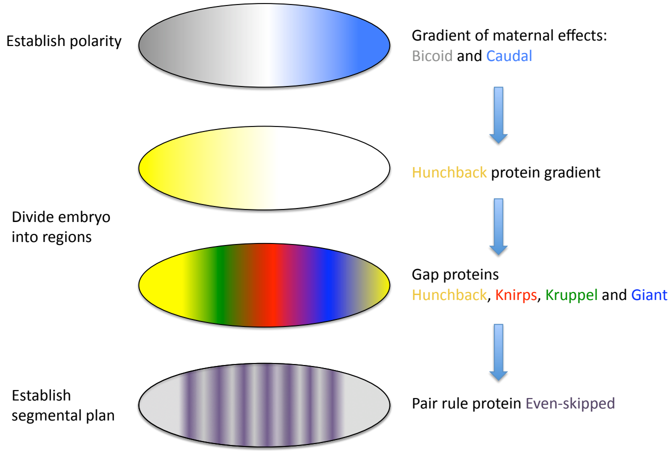
- Model/Simulator
- Evaluator
- Optimiser
    - *trajectory methods*
    - *population-based methods*



Outline

Biological Background - Gene Expression Pattern

- Egg of Drosophila already polarised by maternal effects



Goal of the Model

- Reproducing the expression pattern of the gap genes at Cl. Cyc. 14
- Beginning with expression data at Cl. Cyc. 11
- Experimental data and acquired images comes from the open on-line database FlyEx [1][11]

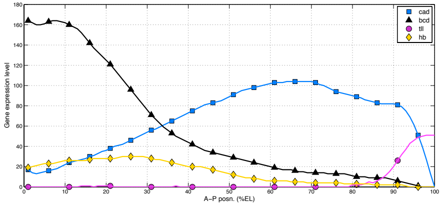


Figure: Quantitative experimental data at cl.cyc.11

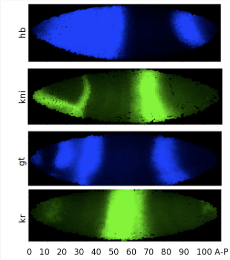


Figure: 2D image at cl.cyc.14

[1] http://flyex.ams.sunysb.edu/flyex/index.jsp

Model of the Cellular-System

- Each compartment is a cell that hosts chemical reactions

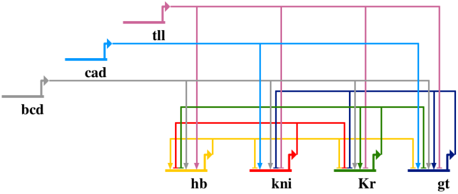
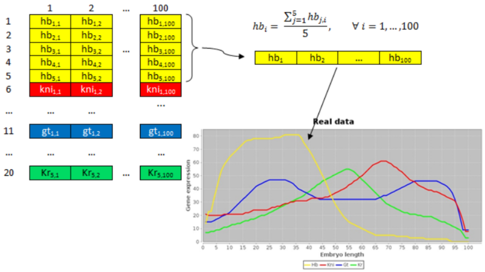


Figure: Intracellular network from literature [10]

- The system size is 10x100
  - y corresponds to the central portion of D-V axis 45%-55%
  - x corresponds to the 0%-100% of the A-P axis
- Grid is fixed
- Hb, Kr, Kni and Gt are able to diffuse

The Parameter Tuning

1. Pre-processing the simulator output

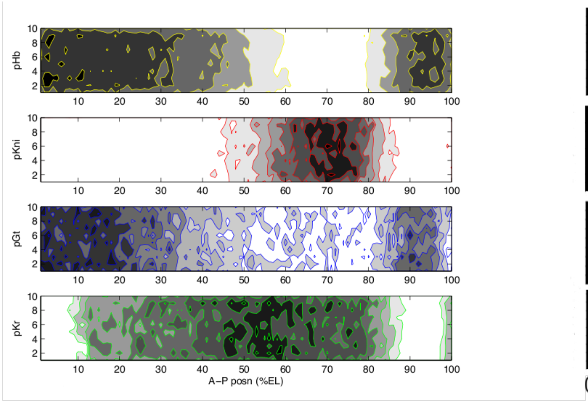


2. Computing the total error as

$$E_{TOT} = \sum_{j=1}^{4} \sqrt{\sum_{i=1}^{100} (o_{j,i} - t_{j,i})^2}$$

- where O is the elaborated simulator output with elements $o_{j,i}$
- where T is the target matrix with elements $t_{j,i}$

Qualitative Results [9, 8]



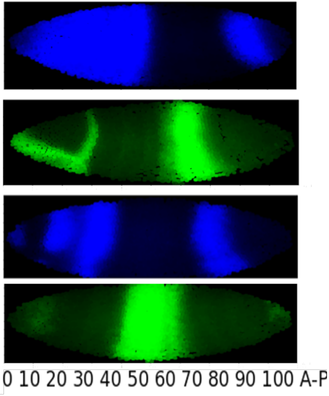


Figure: Simulation results for the four gap genes *hb, kni, gt, Kr* at a simulation time equivalent to the eighth time step of Cleavage Cycle 14A (left) and the corresponding experimental data (right)—% A-P length on the *x* and % D-V width on the *y*

Quantitative Results

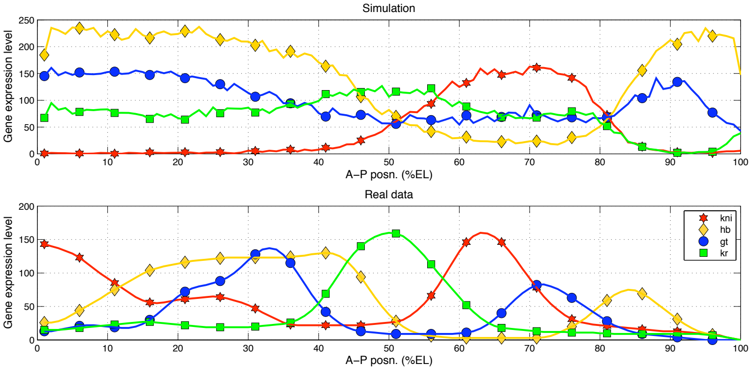


Figure: Quantitative MS-BioNET simulation results for the four gap genes *hb*, *kni*, *gt*, *Kr* at a simulation time equivalent to the eighth time step of cleavage cycle 14A (top) and the corresponding experimental data (bottom)

Outline

Available Theses (in Italian)

http://www.apice.unibo.it/xwiki/bin/view/Theses/Available

Tesi con ALCHEMIST

1. Un linguaggio di alto livello per la descrizione di ecosistemi di servizi pervasivi
2. Interfacce di input e reporting per la simulazione di ecosistemi di servizi pervasivi

Outline

Stefania Bandini, Sara Manzoni, and Giuseppe Vizzari.
Crowd Behavior Modeling: From Cellular Automata to Multi-Agent Systems.
In Adelinde M. Uhrmacher and Danny Weyns, editors, *Multi-Agent Systems: Simulation and Applications*, Computational Analysis, Synthesis, and Design of Dynamic Systems, chapter 13, pages 389–418. CRC Press, June 2009.

Federica Ciocchetta and Maria Luisa Guerriero.
Modelling biological compartments in Bio-PEPA.
Electronic Notes in Theoretical Computer Science, 227:77–95, 2009.

Lorenzo Dematté, Corrado Priami, Alessandro Romanel, and Orkun Soyer.
Evolving blenx programs to simulate the evolution of biological networks.
Theoretical Computer Science, 408(1):83–96, 2008.

Michael A. Gibson and Jehoshua Bruck.
Efficient exact stochastic simulation of chemical systems with many species and many channels.
J. Phys. Chem. A, 104:1876–1889, 2000.

Daniel T. Gillespie.
Exact stochastic simulation of coupled chemical reactions.
The Journal of Physical Chemistry, 81(25):2340–2361, 1977.

Sara Montagna, Nicola Donati, and Andrea Omicini.
An agent-based model for the pattern formation in Drosophila Melanogaster.
In Harold Fellermann, Mark Dörr, Martin M. Hanczyc, Lone Ladegaard Laursen, Sarah Maurer, Daniel Merkle, Pierre-Alain Monnard, Kasper Stoy, and Steen Rasmussen, editors, *Artificial Life XII*, chapter 21, pages 110–117. The MIT Press, Cambridge, MA, USA, 2010.
Proceedings of the 12th International Conference on the Synthesis and Simulation of Living Systems, 19-23 August 2010, Odense, Denmark.

Sara Montagna and Andrea Roli.
Parameter tuning of a stochastic biological simulator by metaheuristics.
In Roberto Serra and Rita Cucchiara, editors, *AI*IA 2009: Emergent Perspectives in Artificial Intelligence – XIth International Conference of the Italian Association for Artificial Intelligence Reggio Emilia, Italy, December 9-12, 2009 Proceedings*, volume 5883 of *Lecture Notes in Computer Science*, pages 466–475. Springer Berlin / Heidelberg, 2009.

Sara Montagna and Mirko Viroli.
A computational framework for multi-level morphologies.
In René Doursat, Hiroki Sayama, and Olivier Michel, editors, *Morphogenetic Engineering: Toward Programmable Complex Systems*, "Studies on Complexity" Series. Springer/NECSI, 2010.

Sara Montagna and Mirko Viroli.
A framework for modelling and simulating networks of cells.
Electronic Notes in Theoretical Computer Science, 268:115–129, 2010.
Proceedings of the 1st International Workshop on Interactions between Computer Science and Biology (CS2Bio'10).

Theodore J Perkins, Johannes Jaeger, John Reinitz, and Leon Glass.
Reverse engineering the gap gene network of *Drosophila Melanogaster*.
PLoS Comput Biol, 2(5):e51, 05 2006.

Andrei Pisarev, Ekaterina Poustelnikova, Maria Samsonova, and John Reinitz.
Flyex, the quantitative atlas on segmentation gene expression at cellular resolution.
Nucleic Acids Research, 37(Database-Issue):560–566, 2009.

Repast Development Team.
`http://repast.sourceforge.net/`.
Repast home page.

Swarm Development Team.
`http://www.swarm.org/index.php/Main_Page`.
Swarm home page.

George Mason University.
`http://www.cs.gmu.edu/~eclab/projects/mason/`.
MASON home page.

- Mirko Viroli, Matteo Casadei, Sara Montagna, and Franco Zambonelli.
  Spatial coordination of pervasive services through chemical-inspired tuple spaces.
  ACM Transactions on Autonomous and Adaptive Systems, 2010.

- Mirko Viroli, Franco Zambonelli, Matteo Casadei, and Sara Montagna.
  A biochemical metaphor for developing eternally adaptive service ecosystems.
  In Sung Y. Shin, Sascha Ossowski, Ronaldo Menezes, and Mirko Viroli, editors, *24th Annual ACM Symposium on Applied Computing (SAC 2009)*, volume II, pages 1221–1222, Honolulu, Hawai'i, USA, 8–12 March 2009. ACM.

- Uri Wilensky and CCL.
  `http://ccl.northwestern.edu/netlogo/index.shtml`.
  NetLogo home page.

Bio-ICT Convergence: Filling the Gap Between Computer Science and Biology

Sara Montagna

sara.montagna@unibo.it

Ingegneria Due
Alma Mater Studiorum—Università di Bologna a Cesena

Academic Year 2010/2011