NII Shonan Meeting Report

No. 2011-7

International Symposium on Symbolic Systems Biology

Katsuhisa Horimoto
Katsumi Inoue
Oliver Ray

September 13–17, 2011

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Table of Contents

1 Abstract

This document is the official record of the first International Symposium on Symbolic Systems Biology (ISSSB'11) held between the 13^{th} and 17^{th} of November, 2011, in Shonan, Japan. The report includes a short introduction to the field of Symbolic Systems Biology (SSB), the aims of the meeting, abstracts of the 6 keynote talks, 20 technical presentations (comprising 14 long talks, and 6 spotlight presentations), and an informal analysis of data from questionnaires returned by the participants. Slides of the presentations can be found on the symposium webpage: `http://www.cs.bris.ac.uk/~oray/ISSSB11/`

2 Introduction

Symbolic Systems Biology is a growing area of research involving the application of formal logic-based methods to systems biology and bioinformatics. With biological data being acquired at ever increasing rates, symbolic approaches are being used more and more, in conjunction with numeric techniques, to help formalise expert knowledge and integrate information across different levels of biological abstraction. Recently, a number of symbolic approaches have been developed and usefully applied to a variety of biological problems. Such methods include

- formal logics (e.g., propositional/first-order/modal frameworks)
- computational logics (e.g., constraint/logic/answer-set programs)
- graphical models (e.g., Boolean/Bayesian/Petri nets)
- synthetic inference (abduction/induction)
- formal methods (e.g., model checking/pi-calculus/hybrid logic)
- qualitative reasoning
- action languages
- statistical relational learning

While many of these techniques are discussed in various texts on systems biology (see the useful links below), there has been very little inter-comparison these methods and their respective strengths and weaknesses. Instead we find a toolbox of isolated approaches with no conceptual foundations that allow any relationships to be theoretically explored and practically harnessed. Such a study is urgently needed to help to facilitate research in symbolic systems biology by providing a roadmap of which systems are best suited to which problems and allowing more effective exploitation and re-use of algorithms and data.

At the same time, there is also a need for more collaboration between numerical and symbolic biologists. This is necessary to better understand the advantages and drawbacks of the quantitative and qualitative approaches and progress towards a synergistic integration of the two. Ideally this should be done with advice from experimental biologists who are better informed of hot applications and emerging methods of data acquisition. This will help to provide a real context in which symbolic systems biology can be more usefully developed.

3 Aims

The primary aim is to initiate a systematic comparison of symbolic methods and their biological applications to better understand their strengths, weaknesses and supported features. For example, most methods operate upon networks which can be classified functionally (metabolic pathways, protein interactions, signal transduction, gene regulation, etc.) or formally (synchronous/asynchronous, discrete/continuous, deterministic/probabilistic, cyclic/acyclic, with/without feedbacks, etc.). We believe such features will help to construct a roadmap of what been achieved so far and suggest future directions for further research.

A secondary aim is to identify ways in which purely symbolic methods have been or could be combined with numerical techniques and applied to emerging problems in experimental and synthetic biology. We believe it is important to promote a dialogue between qualitative, quantitative, and experimental biologists. For example, it seems symbolic biologists can benefit from a better understanding of flux-based analyses used by numerical biologists; while numerical biologists can benefit from a better understanding of graph-based methods used by symbolic biologists. We believe both views are necessary to better handle real world noise and uncertainty.

Ideally, we will end up constructing a web site which provides an overview and comparison of the above-mentioned methods and applications. We would also hope to begin a discussion (possibly the topic of a future follow-on meeting) on how such methods could employ a standard data format (SBML/SGML) to allow a common specification language for such problems to facilitate inter-operability of the methods. We would also hope to produce some small and larger exemplars that could be used to compare the features offered by the various approaches and estimate their scalability to realistically sized problems. In this way we see the symposium as the first step towards providing a theoretical footing for the field of symbolic systems biology.

4 Format

The symposium is hosted by the National Institute of Informatics (NII) in Japan as part of its new series of Shonan meetings which provide a premier Asian location for informatics seminars following the successful European Dagstuhl format. These meetings aim to foster discussion of research and exchange of knowledge between world-class scientists, promising young researchers, and practitioners. They are held in the Shonan Village Center (near Tokyo) which offers a combination of facilities for conferences, training, and lodging in a resort-like setting (with a direct train connection from Narita airport).

We intend to encourage participation from a mix of symbolic biologists, numerical biologists, and experimental biologists. We will encourage these experts to explain their approach, demonstrate their tools, and participate in group discussions aimed at comparing the advantages and disadvantages of each approach and exploring the ways of integrating symbolic and numerical techniques.

5 Keynote Talks (Abstracts)

Formal Cell Biology in BIOCHAM

François Fages

INRIA Paris-Rocquencourt

Joint work with: Sylvain Soliman

Models of cellular processes are developed in systems biology with two somewhat contradictory perspectives: on the one hand, for aggregating knowledge on a given process, by building models as much detailed as possible, on the other hand, for making predictions, by building models as much simple as possible but sufficient for answering a given question on a process. These two perspectives can be reconciled by relating formalisms and models by abstraction/refinement relationships and by organizing them in hierarchies of models. We shall report on our experience in this field with the development of the Biochemical Abstract Machine (http://contraintes.inria.fr/biocham) and the use of this modeling environment in cell signaling, cell cycle and cancer therapy optimization.

Multi-Layer modeling in systems biology

Sriram Iyengar

University of Texas

Biological processes and entities are often mediated by multiple sub-processes and entities. Since perturbations in one can affect the functioning of others and of the entity as a whole, it is important to model not only the individual processes but also to understand how these processes interact with each other Clearly, developing models of such interacting entities is a complex task. In this paper we apply computing principles of abstraction and modularization to propose a multi-layer framework for such purposes. An example of such a framework, applied to biological cells, is presented.

Robot Scientists for Biology

Ross King

Aberystwyth University

A Robot Scientist is a physically implemented robotic system that applies techniques from artificial intelligence to execute cycles of automated scientific experimentation. A Robot Scientist can automatically execute cycles of: hypothesis formation, selection of efficient experiments to discriminate between hypotheses, execution of experiments using laboratory automation equipment, and analysis of results. We first developed the Robot Scientist 'Adam' to investigate yeast (Saccharomyces cerevisiae) functional genomics. Adam has autonomously identified genes encoding locally 'orphan' enzymes in yeast. This is the first time a machine has discovered novel scientific knowledge. We have now developed the Robot Scientist 'Eve' to automate and integrate: drug screening, hit conformation, and QSAR development. Eve utilises novel synthetic biology screens that combine the advantages of target-based and cell-based assays. This combination of novel automation, active machine learning, and synthetic biology assays enables much faster and cheaper screening. Our focus is on neglected tropical diseases, and with in the last six months Eve has successfully confirmed hits against three targets in six different parasites.

Automated Discovery of Food Webs from Ecological Data using Logic-based Machine Learning

Stephen Muggleton

Imperial College London

Joint work with: David A. Bohan, Alan Raybould, Alireza Tamaddoni-Nezhad

Networks of trophic links (food webs) are used to describe and understand mechanistic routes for translocation of energy (biomass) between species. However, a relatively low proportion of ecosystems have been studied using food web approaches due to difficulties in collecting and verifying observations on large numbers of trophically-interacting species. In this paper we demonstrate that Machine Learning of food webs, using a logic-based approach called Abductive ILP (A/ILP), can generate plausible and testable food webs from empirical, field sample data. Our example training data come from a national-scale Vortis suction sampling of invertebrates from arable agricultural fields in the United Kingdom. We found that 45 invertebrate species or taxa, representing approximately 25% of the species or taxa in the sample and about 74% of the invertebrate individuals included in the learning, were hypothesized to be linked. Generalist

and omnivorous carabid beetles were hypothesized to be the dominant predators of the system. We were, however, surprised by the importance of carabid larvae suggested by the machine learning as predators of a wide variety of prey. High probability links were hypothesized for widespread, potentially destabilizing, intra-guild predation; predictions that could be experimentally tested. Many of the high probability links in the model have already been observed or suggested for this system supporting our contention that A/ILP learning can produce plausible food webs from sample data, independent of our preconceptions about who eats whom. In particular, well-characterised links in the literature correspond with links ascribed with high probability through A/ILP. We believe that this very general Machine Learning approach has great power and could be used to extend and test our current theories of agricultural ecosystem dynamics and function.

Programming Cells

Andrew Phillips

Microsoft Research

Living cells are highly sophisticated computational machines, constantly processing information to survive, grow and reproduce. The software that allows a cell to function is stored inside the cell as DNA, which codes for proteins that ultimately determine the cell's behaviour. If we could program cells as effectively as we program digital computers, we could address some of the key challenges facing society in areas of health, food and energy production. In developing this technology we could also gain insight into the workings of life itself. In spite of this promise, many fundamental challenges still lie ahead, including how to design and implement cell programs, and how to ensure that they behave as intended. To overcome these challenges will require programming languages and compilers that can take a high-level description of what the cell should do and compile this to low-level DNA machine code. In this talk I will describe some of our initial efforts in developing computer software for programming cells. The programmer writes a high-level description of the desired behaviour of the cell, and the software is used to simulate and analyse this behaviour, and to automatically generate the corresponding DNA code. I will also present some of the software we have been developing for programming DNA circuits that perform logical computations. Such circuits could eventually be inserted into cells to monitor the health of the cell and compute appropriate responses.

Symbolic Systems Biology: a Perspective

Carolyn Talcott

SRI International

The talk will begin by saying what we mean by "Symbolic Systems Biology" (SSB). This will be followed by an overview of the Pathway Logic work, as an example of SSB. The talk will conclude with a discussion of opportunities for synergy amongst different SSB efforts and challenges, including representation and semantics, and combining multiple views into a bigger picture.

6 Technical Presentations (Abstracts)

Attractor detection and control of Boolean networks

Tatsuya Akutsu

Kyoto University

We have been studying attractor detection and control problems on Boolean networks. For attractor detection, our main objective is to break the trivial $O(2^n)$ bound (n is the number of nodes), and we developed $O(1.587^n)$ time and $O(1.799^n)$ time algorithms for singleton attractor detection of Boolean networks consisting of AND/OR functions and analyzing functions, respectively. For control problems, we recently proved that control of probabilistic Boolean networks (probabilistic extension of Boolean networks) is Σ_2^p-hard, which suggests that control of probabilistic Boolean networks is harder than control of Boolean networks. We also developed practical integer programming-based methods for solving both attractor detection and control problems.

Gene Essentialities of Bacterial Systems

Tomoya Baba

Transdisciplinary Research Integration Center, National Institute of Genetics

Joint work with: Barry L. Wanner, Masaru Tomita and Hirotada Mori

Bacteria have very simple cell structures and biological systems, however, have been evolutionarily changed by themselves and adapted to their living environments all of the Earth. Escherichia coli K-12, a non-pathogenic enterobacteria living in human gut, is most characterized bacterium in genetics, biochemistry and molecular biology. It has about 4,300 genes meaning "Open Reading Frames (ORFs)" as protein coding on the genome. We successfully constructed single-gene knockout mutants for 3,985 genes that revealed those genes non-essential on the biological system in E. coli cells. On the other hand, 303 genes were unable to be disrupted and resulted as essential genes. Some of single-gene knockout mutants, meaning genotypes, showed phenotypes that could grow normally in LB (nutrient-rich) medium, however, not in glucose minimum (nutrient-pure) medium. Those genotypes were thought to be corresponding to metabolic pathway and/or related genes as conditionally essential ones. I would like to briefly introduce our experimental approaches for bacterial systems biology.

Using text extraction and reasoning to construct pharmaco-kinetic pathways and further reason with them to discover drug-drug interactions

Chitta Baral

Arizona State University

Joint work with: L. Tari, S. Anwar, S. Liang, Jorg Hakenberg, and J. Cai.

Biological pathways are seen as highly critical in our understanding of the mechanism of biological functions. To collect information about pathways, manual curation has been the most popular method. However, pathway annotation is regarded as heavily time- consuming, as it requires expert curators to identify and collect information from different sources. Even with the pieces of biological facts and interactions collected from various sources, curators have to apply their biological knowledge to arrange the acquired interactions in such a way that together they perform a common biological function as a pathway. In this talk, we first discuss an approach for automated pathway synthesis that acquires facts from hand-curated knowledge bases. To comprehend the incompleteness of the knowledge bases, our approach also obtains facts through automated extraction from Medline abstracts. An essential component of our approach is to apply logical reasoning to the acquired facts based on the biological knowledge about pathways. By representing such biological knowledge, the reasoning component is capable of assigning ordering to the acquired facts and interactions that is necessary for pathway synthesis. We demonstrate the feasibility of our approach with the development of a system that synthesizes pharmacokinetic pathways. Next, we will discuss an approach that integrates text mining and automated reasoning to derive drug-drug interactions. Through the extraction of various facts of drug metabolism, not only the drug-drug interactions that are explicitly mentioned in text can be extracted but also the potential interactions that can be inferred by reasoning.

Problem Decomposition for Reasoning with Biological Networks

Gauvain Bourgne

National Institute of Informatics

Joint work with: Katsumi Inoue

Short abstract (less than 200 words): Systems biology typically requires reasoning on huge networks (metabolic pathways, signaling networks, gene regulation...) that can be challenging for centralized solvers. We propose here some decomposition approach to divide the reasoning between loosely coupled subnetworks, and shows some promising first results on accessibility problems in metabolic pathways.

On Construction of Probabilistic Boolean Networks

Wai-Ki Ching

The University of Hong Kong

Modeling genetic networks is an important in problem genomic research. Boolean Network (BN) and its extension Probabilistic Boolean networks (PBN) have been proposed to model genetic regulatory interactions. In a PBN, its steady-state distribution gives very important information about the long-run behavior of the network. The construction of PBNs from a given transition probability matrix and a given set of BNs is an inverse problem of huge size. We propose a maximum entropy approach for the above problem. Newton's method in conjunction with conjugate gradient method is then applied to solving the inverse problem. We investigate the convergence rate of the proposed method. Numerical examples are also given to demonstrate the effectiveness of our proposed algorithm.

Complex PRISM models for analyzing very large biological sequence data, plus a few notes on probabilistic abductive logic programming

Henning Christiansen

Roskilde University

Joint work with:Christian Theil Have, Ole Torp Lassen and Matthieu Petit

We give a short overview of results from our work on analysis of biological sequence data using Sato et al's PRISM system. As sequence analysis may not be the key issue in systems biology, we will focus here on how we tackled the scaling problems that arise when using PRISM for: - very long sequences - complex models with many parameters and dependencies. We did experience PRISM and probabilistic logic programming in general as a very flexible tool for sequence analysis, which (we hypothesize) may also be the case for systems biology where we may expect to face similar scaling problems. Finally, we will present a single slide summing up out work on probabilistic abductive logic programming.

Identifying Candidate Pathways to Explain Phenotypes in Genome-Wide Mutant Screens

Mark Craven

University of Wisconsin

Joint work with: Debbie Chasman

New genome-wide assays are enabling biologists to detect which gene products in host cells are exploited by viruses. In these assays, the expression of each gene in a host cell is suppressed or abolished, and then viral replication is measured to determine the effect of the gene on the virus. Although these methods can indicate which genes have roles in viral replication, they do not elucidate how these genes are organized into biological pathways that mediate host-virus interactions. We are developing computational methods that transform the measurements from these assays into hypotheses that predict which host genes most directly interact with the virus, and indicate the pathways in the cell that relate the other implicated genes to viral replication. We are exploring two alternative approaches to this task: one is based on solving an integer program, and the other is based on doing inference in a Markov network.

Model of Double Strand Break of DNA in Logic-Based Hypothesis Finding

Andrei Doncescu

LAAS-CNRS/University Paul Sabatier Toulouse

Joint work with: Katsumi Inoue and Barthelemy Dworkin

In the past decades a lot of researches have been conducted to investigate genes regulation in order to increase the understanding of the mechanisms responsible for diseases and thus to find new treatments. One strategy to achieve this goal would be to inhibit or to trigger a part of the pathway that produces the disease. A strategy would be to first identify the signaling pathway whose activity contributes to the disease and then look for drugs that inhibit or trigger the pathway acting on its internal components and reactions.

The diversity of methods relevant to symbolic systems biology

Randy Goebel

University of Alberta

Joint work with: Guohui Lin

Symbolic reasoning about systems biology requires exactly what symbolic reasoning requires in any application area: the construction of detailed models, these days only doable by machine learning, and goal directed hypothesis management, such as provided by abductive reasoning systems. We provide motivation for hypothesis management that requires integration of modeling, reasoning and visualization, and note several tough problems for which progress must be made, in order to achieve advantages of symbolic reasoning about systems biology.

Abduction in meta-reasoning

Katsumi Inoue

National Institute of Informatics

This talk is given as a method of "object invention" in the discussion chaired by Stephen Muggleton on the role of "predicate invention" in Systems Biology modeling. I firstly show how the combination of meta-reasoning and abduction is powerful enough to infer missing formulas in general. Then, with an axiom set for causal reasoning allowing transitivity, meta-level abduction can infer hidden causal rules and unknown facts from incomplete knowledge to explain empirical rules as observations (Inoue, Furukawa and Kobayashi, ILP'09). In biological pathways, meta-level abduction can infer missing links and nodes in networks containing both positive and negative causal effects (Inoue, Doncescu and Nabeshima, ILP'10). Case studies are then presented in p53 signaling networks, in which causal relations are abduced to suppress a tumor with a new protein (Tran and Baral, 2009) and to stop DNA synthesis when damage occurs.

Modeling of signaling pathways based on Petri nets

Hiroshi Matsuno

Yamaguchi University

The mechanism of signaling pathway is complex, consisting of distinct reactions such as complex formation, catalytic reaction, and transportation. These reactions combine to propagate 'signals' by changing states of participating molecules 'inactive' to 'active' in sequence from the membrane of a cell to the target site in the nucleus. This means that signaling pathways essentially involve "dynamic elements" working for the signal propagation as well as "static elements" accounting for the states of molecules. In this talk, after presenting the components of these reactions, we demonstrate how to construct the Petri net structure of a signaling pathway, namely, the network of static elements, using an apoptosis pathway. To realize the signal propagation, reaction speeds should be incorporated into the Petri net model, but all of the reaction speed cannot be measured by biological experiments. Then an algorithm is given to estimate the delay times of the Petri net model.

Robustness analysis of yeast cell cycle in silico and in vivo.

Hisao Moriya

Okayama University

Intracellular parameters such as gene expression require optimization, such that cellular functions may be performed effectively. Fluctuations in these parameters lead to various cellular defects. Overexpression of genes involved in proliferation of cancer cells due to gene amplification is a prime example. On the other hand, in order to maintain cellular functions despite environmental change, mutation, and noise in intracellular biochemical reactions, these parameters may have certain permissible ranges, a characteristic termed robustness, which is commonly observed in various cellular systems.

We developed a method designated genetic tug-of-war (gTOW), by which we can measure the limit of gene overexpression in the budding yeast /Saccharomyces cerevisiae/. Using gTOW, we measured the copy number limits of 30 cell-cycle regulators in budding yeast(Moriya et al, 2006). The data was used to reveal the robustness profile of the cell-cycle regulatory system, and to evaluate and refine the integrative mathematical model of the budding yeast cell cycle (Kaizu et al, 2010; Moriya et al, 2006).

Inference of Biological Networks from Experimental Data

Anne Poupon

CNRS

Joint work with: Pauline Gloaguen, Sarah Cohen-Boulakia and Christine Froidevaux

With the rise of high-throughput experimental methods and the rapid development of bioinformatics methods, there has been a dramatic increase in the quantity and heterogeneity of available data in systems biology. There is no method to integrate this information in molecular networks, and it has become impossible for the human mind to make this integration. In the case of genetic networks, machine learning methods have been used to deduce the rules to build new networks from established networks and initial data. However, our knowledge on signalling networks is not sufficient to build the necessary learning set. We have chosen to develop a method that would mimic the scientist's reasoning, but with the computer's integration and data treatment power. We have started the development of a knowledge-based method, built on the consequence-finding system in first-order logic SOLAR and have proved that this system is able to reconstruct signalling pathways.

Biological pathway inference with answer set programming

Oliver Ray

University of Bristol

This talk summarises the results of three investigations into the use of Answer Set Programming (ASP) for biological pathway analysis. The first piece of work (with Ross King and Ken Whelan) was presented at IIBM'10 and shows how ASP can be used to very naturally represent and reason about simple reaction networks. In particular it shows how the stability of ASP models eliminates certain unfounded steady states that have confounded some other logical approaches. The second piece of work (with Katsumi Inoue and Takehide Soh) was presented at ANB'10 and shows how ASP can also be used naturally incorporate numeric parameters into the formalism in order to rank candidate solutions. The third piece of work (with Ross King and Ken Whelan) was presented at ILP'09 and shows how ASP has been successfully used to revise a state-of-the-art metabolic network in order to make it consistent with observational data acquired by a robot scientist.

Some contributions to transcriptomic data analysis and gene regulation learning

Céline Rouveirol

LIPN, Université Paris Nord

Joint work with: M. Elati, L. Létocart, K. Mouhoubi and F. Radvnayi

We will mainly describe two contributions of symbolic approaches to the analysis of transcriptomic data. The first contribution concerns one of the most challenging tasks in the post-genomic era, the reconstruction of transcriptional regulation networks. We have proposed a data mining system for inferring transcriptional regulation relationships from RNA expression values, particularly suitable for the detection of cooperative transcriptional regulation. We model regulatory relationships as labelled two-layer gene regulatory networks, and describe a method for the efficient learning of these bipartite networks from discretized expression data sets. The second contribution is a new heuristic approach based on a graph algorithm for the efficient extraction of itemset patterns in noisy binary contexts. This last method is based on maximal flow/minimal cut algorithms to find dense subgraphs of 1 in the graph associated to the boolean data matrix. Both these methods have been applied to analyse yeast and human gene expression datasets.

Repair and Prediction (under Inconsistency) in Large Biological Networks with Answer Set Programming

Torsten Schaub

University of Potsdam

Joint work with: Martin Gebser, Carito Guziolowski, Anne Siegel, Sven Thiele, Philippe Veber

We address the problem of repairing large-scale biological networks and corresponding yet often discrepant measurements in order to predict unobserved variations. To this end, we propose a range of different operations for altering experimental data and/or a biological network in order to re-establish their mutual consistency—an indispensable prerequisite for automated prediction. For accomplishing repair and prediction, we take advantage of the distinguished modeling and reasoning capacities of Answer Set Programming. We validate our framework by an empirical study on the widely investigated organism *Escherichia coli*.

Predicting Gene Knockout Effects by Minimal Pathway Enumeration

Takehide Soh

Transdisciplinary Research Integration Center

Joint work with: Katsumi Inoue, Tomoya Baba and Toyoyuki Takada

It is an important subject to analyze how gene knockouts affect the phenotype of organisms. Although breeding knockout organisms is one way, it is costly. In this talk, we show a method to predict gene knockout effects on Escherichia coli (E. coli) utilizing biological databases such as KEGG and EcoCyc, which have collected biological knowledge and experimental results. We construct biological networks from them and represent its causal relations in propositional formulas. We then execute a model generator and enumerate minimal active pathways, which are minimal subsets of a given biological network using source metabolites to produce target metabolites. We simulate the effects of gene knockouts by measuring the difference of minimal active pathways between original networks and knockout ones. We apply this method to predict the gene knockout effects of E. coli and have comparisons with the growth rates of every single gene knockout strain, which are obtained from biological experiments.

Comparative analysis of liver gene expression profiles in mouse C57BL/6J and MSM/Ms strains

Toyoyuki Takada

National Institute of Genetics

Since fine genome sequence of a standard classical laboratory mouse strain C57BL/6J (B6) and information of genetic variation such as SNPs of many inbred strains are now available, it is becoming realistic to explore the genome functions based on statistical linking of phenotypes and genome diversity of different inbred strains. An inbred strain MSM/Ms is derived from Japanese wild mice, Mus musculus molossinus. The genome sequence of MSM/Ms is divergent from that of B6, whose genome is predominantly derived from Western European wild mouse, Mus musculus domesticus. MSM/Ms exhibits a number of quantitative complex traits markedly different from those of B6 such as body growth and energy metabolism. We are conducting microarray based gene expression analysis of metabolism-related tissues to try and find which genes and/or genetic networks are responsible for these phenotypic differences. In this meeting, we will present preliminary data analysis of liver gene expression profiles in B6 and MSM/Ms strains.

Predicting essential genes via impact degree on metabolic networks

Takeyuki Tamura

Kyoto University

The impact degree is a measure of the robustness of a metabolic network against deletion of single or multiple reaction(s). Although such a measure is useful for mining important enzymes/genes, it was defined only for networks without cycles. In this work, we extend the impact degree for metabolic networks containing cycles and develop a simple algorithm to calculate the impact degree. Furthermore we improve this algorithm to reduce computation time for the impact degree by deletions of multiple reactions. As a result of preliminary experiments, 12 of the top 14 genes associated with high impact degree were included in the list of essential genes of Baba et al.

Understanding Intracellular Signalling in Bacterial Chemotaxis

Marcus Tindall

University of Reading

Joint work with: Prof J.P. Armitage

For nearly 40 years experimental and theoretical researchers have worked together to understand how bacterial chemotactic species, in particular Escherichia coli, function. In this presentation I will provide recent examples of how we have formulated mathematical models of the intracellular signalling networks within E. coli and R. sphaeroides to understand how the biochemical signalling cascades within these bacteria function and affect their response to external stimuli.

7 Analysis of Participant Questionnaires

In advance of the symposium we asked all 31 participants to fill out a questionnaire which included the following questions concerning their target problems and methodologies:

```
3. What is your specialty?
   - Mathematics (incl. logic)
   - Computer science (incl. AI, programming, algorithms)
   - Bioinformatics (incl. systems biology)
   - Biology
   - others (please specify)

4. What kind of biological problems are you solving?
   (multiple answers allowed)
   - metabolic pathways
   - protein-protein interaction
   - signaling networks
   - gene regulatory networks
   - sequence analysis
   - visualization
   - evolution
   - ecology
   - modeling (of what?)
   - others (please specify)

5. Please choose your research interests and methodologies from
the following list (multiple answers allowed).

 (a) formal logics
   (a1) propositional/Boolean logic
   (a2) first-order logic
   (a3) higher-order logic
   (a4) modal or temporal logic
   (a5) fuzzy logic
   (a6) others (please specify)

 (b) computational logics
   (b1) constraint programming
   (b2) logic/answer-set programming
   (b3) action languages
   (b4) Boolean satisfiability/SMT
   (b5) multi-agent systems
   (b6) others (please specify)

 (c) graphical models
```

```
  (c1) Boolean networks
  (c2) Bayesian networks
  (c3) Petri nets
  (c4) cellular automata
  (c5) graph theory
  (c6) others (please specify)

(d) synthetic inference
  (d1) abduction
  (d2) induction
  (d3) analogical/case-based reasoning
  (d4) planning
  (d5) others (please specify)

(e) formal analytic methods
  (e1) model checking
  (e2) pi-calculus
  (e3) hybrid logic
  (e4) qualitative reasoning
  (e5) others (please specify)

(f) machine learning methods
  (f1) classification
  (f2) relational learning
  (f3) probabilistic/statistical learning
  (f4) neural networks
  (f5) others (please specify)

(g) control and numeric methods
  (g1) differential equations
  (g2) linear/non-linear dynamical systems
  (g3) stochastic processes
  (g4) optimization techniques
  (g5) others (please specify)

(h) others: (please specify)
```

From the answers of the participants, we summarize the following statistics. Figure 1 shows specialities of participants. As the figure shows, almost all participants come from the domain of Computer Science and Bioinformatics. Figure 2 shows clustering of participants according to their research interests. Figure 3 shows a bipartite graph $(N_1 \cup N_2, E)$ representation of the answers of the questionnaire, where N_1 is a set of biological problems that participants are solving, N_2 is a set of methods and E is a set of edges. Each edge width corresponds to the number of participants who checked both $n \in N_1$ and $n' \in N_2$. To emphasize major relations, we omit the edges corresponding to numbers less than six.

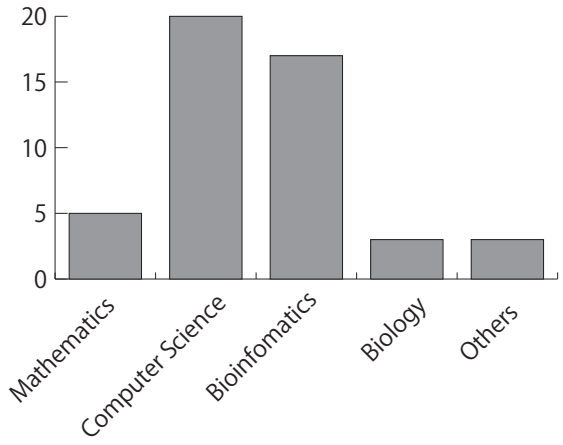

Fig. 1. Speciality of Participants

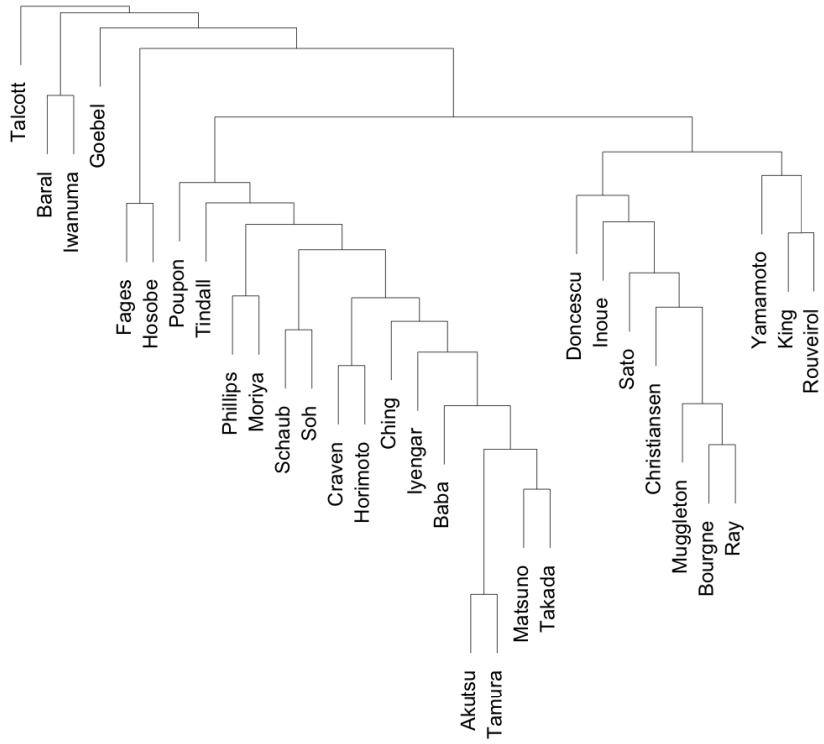

Fig. 2. Clustering of Participants according to their Research Interests

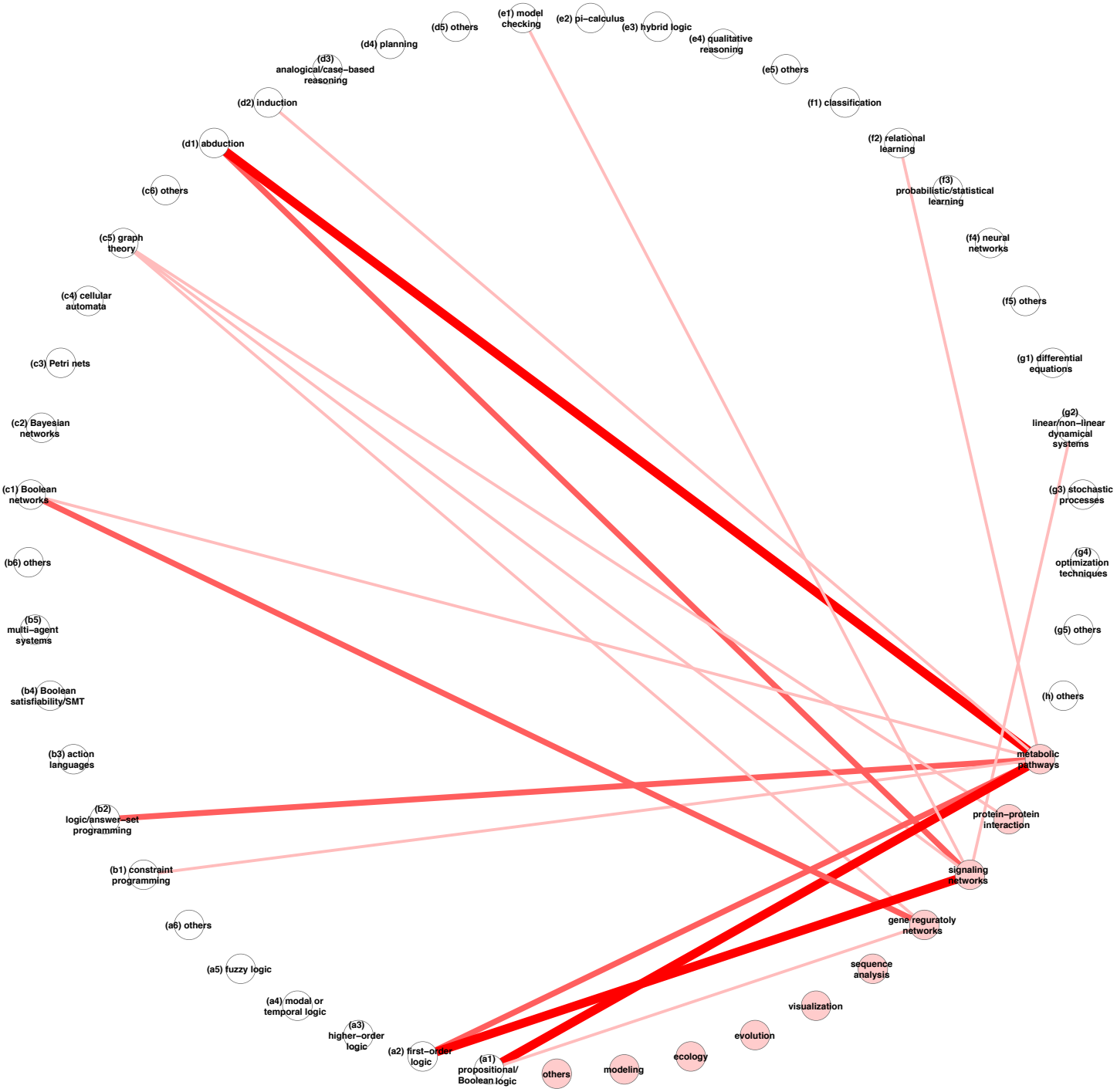

Fig. 3. Relation between Target Problems and Methodology

8 Programme

Sunday 13th November

15.00 Check In

19.00 Reception

Monday 14th November

07.30 Breakfast

08.55 Opening Address: **Oliver Ray**

09.00 Keynote: **Carolyn Talcott**
Symbolic Systems Biology: a Perspective

09.45 Agenda Discussion: **Katsumi Inoue & Oliver Ray**

10.30 Coffee

11.00 Talk: **Chitta Baral**
Using text extraction and reasoning to construct pharmaco-kinetic pathways and further reason with them to discover drug-drug interactions

11.30 Talk: **Hiroshi Matsuno**
Modelling of signalling pathways based on Petri nets

12.00 Lunch

13.30 Group Photo

14.00 Talk: **Torsten Schaub**
Repair and Prediction (under Inconsistency) in Large Biological Networks with Answer Set Programming

14.30 Talk: **Anne Poupon**
Inference of Biological Networks from Experimental Data

15.00 Talk: **Marcus Tindall**
Understanding Intracellular Signalling in Bacterial Chemotaxis

15.30 Coffee

16.00 Keynote: **Francois Fages**
Formal Cell Biology in BIOCHAM

16.45 Talk: **Randy Goebel**
The diversity of methods relevant to symbolic systems biology

17.15 Discussion (Methods): **Carolyn Talcott**

17.45 Free Time

18.00 Dinner

Tuesday 15th November

07.30 Breakfast

09.00 Keynote: **Sriram Iyengar**
Multi-layer modelling in systems biology

09.45 Talk: **Celine Rouveirol**
Some contributions to transcriptomic data analysis and gene regulation learning

10.15 Spotlight: **Andrei Doncescu**
Model of Double Strand Break of DNA in Logic-Based Hypothesis Finding

10.30 Coffee

11.00 Talk: **Tatsuya Akutsu**
Attractor detection and control of Boolean networks

11.30 Discussion (Methods): **Andrew Phillips**

12.00 Lunch

13.30 Talk: **Marc Craven**
Identifying Candidate Pathways to Explain Phenotypes in Genome-Wide Mutant Screens

14.00 Talk: **Hisao Moriya**
Robustness analysis of yeast cell cycle in silico and in vivo

14.30 Spotlight: **Toyoyuki Takada**
Comparative analysis of liver gene expression profiles in mouse C57BL/6J and MSM/Ms strainss

14.45 Discussion (Genes): **Marcus Tindall**

15.30 Coffee

16.00 Keynote: **Ross King**
Robot Scientists for Biology

16.45 Spotlight: **Wai-Ki Ching**
On Construction of Probabilistic Boolean Networks

17.00 Free Time

18.00 Dinner

19.00 Discussion (Logic): **Stephen Muggleton**

19.30 Talk: **Oliver Ray**
Biological pathway inference with answer set programming

20.00 Talk: **Katsumi Inoue**
Abduction in meta-reasoning

Wednesday 16th November

07.30 Breakfast

09.00 Keynote: **Andrew Phillips**
Programming Cells

09.45 Spotlight: **Gauvain Bourgne**
Decomposition for Reasoning with Biological Networks

10.00 Spotlight: **Tomoya Baba**
Gene essentialities of bacterial systems

10.15 Spotlight: **Takehide Soh**
Predicting Gene Knockout Effects by Minimal Pathway Enumeration

10.30 Coffee

11.00 Talk: **Takeyuki Tamura**
Predicting essential genes via impact degree on metabolic networks

11.30 Discussion (Networks): **Katsuhisa Horimoto**

12.00 Lunch

13.30 Excursion (Kamakura)

18.00 Banquet (Kamakura)

20.30 Free Time

Thursday 17th November

07.30 Breakfast

07.30 Check Out

09.15 Keynote: **Stephen Muggleton**
Automated Discovery of Food Webs from Ecological Data using Logic-based Machine Learning

10.00 Talk: **Henning Christiansen**
Complex PRISM models for analyzing very large biological sequence data, plus a few notes on probabilistic abductive logic programming

10.30 Coffee

11.00 Closing Discussion: **Oliver Ray & Katsumi Inoue**

Acknowledgements

The organisers thank Takehide Soh for his assistance with local organisation, compilation of this report, and analysis of the participant questionnaires. We also thank for Junji Oshima and Ayako Sumi for their cordial support.

List of Participants

First name	Last name	Institution
Tatsuya	AKUTSU	Kyoto University
Tomoya	BABA	Transdisciplinary Research Integration Center
Chitta	BARAL	Arizona State University
Gauvain	BOURGNE	National Institute of Informatics
Wai-ki	CHING	The University of Hong Kong
Henning	CHRISTIANSEN	Roskilde University
Mark	CRAVEN	University of Wisconsin
Andrei	DONCESCU	LAAS-CNRS
François	FAGES	INRIA, Paris-Rocquencourt
Randolph	GOEBEL	University of Alberta
Katsuhisa	HORIMOTO	National Institute of Advanced Industrial Science and Technology
Hiroshi	HOSOBE	National Institute of Informatics
Katsumi	INOUE	National Institute of Informatics
Koji	IWANUMA	University of Yamanashi
Sriram	IYENGAR	University of Texas
Ross	KING	Aberystwyth University
Hiroshi	MATSUNO	Yamaguchi University
Hisao	MORIYA	Okayama University
Stephen	MUGGLETON	Imperial College London
Andrew	PHILLIPS	Microsoft Research
Anne	POUPON	CNRS-INRA
Oliver	RAY	University of Bristol
Céline	ROUVEIROL	LIPN, University Paris Nord
Taisuke	SATO	Tokyo Institute of Technology
Torsten	SCHAUB	University of Potsdam
Takehide	SOH	Transdisciplinary Research Integration Center
Toyoyuki	TAKADA	National Institute of Genetics
Carolyn	TALCOTT	SRI International
Takeyuki	TAMURA	Kyoto University
Marcus	TINDALL	University of Reading
Yoshitaka	YAMAMOTO	University of Yamanashi