Computational investigation of dynamic response of small networks: a research proposal

Rosemary Braun
National Cancer Institute
NIH, Bethesda, MD
braunr@mail.nih.gov

Marek Kwiatkowski*
School of Informatics
University of Edinburgh
M.Kwiatkowski@ed.ac.uk

Alexander Mikheyev
Section of Integrative Biology
University of Texas
mikheyev@mail.utexas.edu

August 2009

Abstract We propose to determine the reaction of every network of 3-5 nodes to a standardised signal. We see the networks as cellular signalling modules, but other interpretations may apply as well. We argue that the results of this project may shed light on evolutionary properties of small molecular networks. The computational framework we propose to use is *probabilistic model checking*, which guarantees a mathematically sound and completely automatic classification of responses. We invite the reader to undertake the research, with our assistance if required.

1 Problem

Imagine a network with very few nodes, say 3 to 6. Every node may be connected with any other with a "positive" edge, a "negative" edge or no edge at all. Furthermore, every node contains a very small population (order of 10) of smaller entities, each of which can be either "on" (or active) or "off" (inactive). We can equip this simple structure with dynamics in the following way: if a node, say A, is connected with another, say B, by a positive edge, then an active entity inside A may pass this status to an inactive entity in B. Similarly, if the edge is of the negative kind, then an active entity inside A can inactivate an active one in B. We may also allow an inactive entity to spontaneously become active and/or vice versa.

The primary interpretation of this model is intracellular signalling. Nodes are protein kinds and the small entities are actual protein molecules which can be either active or not. The edges denote the effect that the protein of one kind may have on one of another, namely activation (phosphorylation) or deactivation (dephosphorylation). The rest of this document is based on this interpretation, but we shall keep everything general enough for the setting and results to be applied to another problem domain if one is identified.

Since the networks are interpreted as signal transduction modules, we need to identify the source and destination of the signal. In each network, therefore, we shall arbitrary choose one node to serve as the Receptor protein type (i.e. source) and another one to be the Effector species (i.e. destination, or sink). We call every possible configuration of n nodes with Effector and Receptor an *n-topology*, and the central question we are interested in is: **How does the dynamic behaviour of Effector proteins in re-**

*to whom correspondence should be directed.

sponse to an standardised external stimulation of Receptor proteins depend on the configuration of the network?

There exists a study [6] which addresses precisely this question, and which has been the inspiration for this proposal. In this paper, the dynamic responses of 3-, 4- and (partially) 5-topologies are described by setting up an ODE model for each topology and assigning the response to one of four response classes (constant, Gaussian, switch and oscillatory). Perhaps the most interesting finding is the significant increase of oscillatory topologies when individual proteins are allowed to sponaneously change their state from on to off and back.

We propose to repeat this experiment, but with one crucial change: we insist on using stochastic dynamics instead of ODEs. It is well known that there exist biochemical networks for which the naïve ODE dynamics give erroneous results (e.g. Repressilator [3]). This is always linked to low copy numbers of molecules, which in turn is relatively common in signalling. By using stochastic dynamics we ensure that the dynamic behaviour we obtain is indeed the one that would be observed in nature, and also make the results amenable to a very specific kind of analysis as described below (see Sec. 2.1 Multistability).

In the remainder of this document we further describe the kind of questions we are most interested in answering (Sec. 2 Goals) and present the computational framework we think should be used (Sec. 3 Methods).

2 Goals

We suggest that the analysis of the results should focus on two notions: multistability and robustness. We briefly discuss the meaning and importance of these concepts below.

2.1 Multistability

A stochastic network is *multistable* (*bistable*) if it admits many (two) substantially different dynamic behaviours. In our setting this means that the responses of the given topology differ between different runs (simulation) to the extent that they fall into different response classes.

At the level of populations, multistability gives rise to a situation where in a constant environment phenotypical diversity exists without a corresponding genotypical variation: the same biochemical network (genotype) expressed across the population produces a range of responses (phenotypes). Apart from being an unorthodox way of achieving phenotypical variation, such mechanism is also interesting because it may evolve through group selection, since a group of genetically identical individuals exhibiting a range of phenotypes is more robust against environmental pressure that one where the phenotype is constant.

The questions about multistability that could be answered by the research we propose are: What are the multistable architectures? How many of them are there? How are they distributed in the space of all topologies? Is multistability in any way robust against simple changes in the network topology? What is the minimum size of a topology that is multistable? We remark that neither of these questions could be addressed by the study [6], because ODE models are unable to recognise multistability.

2.2 Robustness

By *mutational robustness* – henceforth simply *robustness* – we mean the extent to which a given phenotype persists despite changes to the underlying genetic architecture. To talk about robustness of small dynamics networks we suggest to use the perspective of A. Wagner [7]. In this setting, we need two ingredients to quantify robustness: the space of genotypes, i.e. a set with a binary relation specifying which genotype can mutate into which; and a genotype-phenotype map that tells us when two genotypes give the same phenotype. Robustness of a phenotype can then be defined as the number of genotypes coding for it.

In our case, the space of genotypes is the space of all topologies of a given size, and we mandate that one topology can mutate into another if they differ by only one edge. The phenotype is of course the dynamic response class, which will have been computed for every topology. The questions to address here are:

which phenotype is the most robust? Can every phenotype be reached in one step from any other? What are the relative sizes of the *neutral networks*, i.e. the subspaces of genotypes that translate to the same phenotype? Does robustness promote *evolvability*, as argued in [7]?

3 Methods

There are many ways of studying dynamics of stochastic networks. In our case, the key challenge is to have a response classification method that is fully automatic, because of the number of topologies we want to look at. The computational cost is also a factor, for the same reason, but luckily the task parallelises very well.

We propose to encode every topology as a Continuous Time Markov Chain (CTMC) [4] and to specify response classes as logical formulae. Because an individual response is essentially a time series – or a simulation trace – the logic to use is Linear Temporal Logic (LTL) [2]. An LTL formula is either true or false of a given trace and what we essentially look for is its validity for a trace typical of a given topology.

The computational technique that allows us to quantitatively answer such questions is called *probabilistic model checking* [2, 5]. Given an LTL formula and a CTMC, a probabilistic model checker such as PRISM [1] computes the probability that a single execution of the chain satisfies the formula or, in other words, the fraction of accepted simulation traces. Having a number associated with every response class for every topology is extremely useful, because it makes it possible to distinguish true multistability from occasional stochastic fluctuations of the dynamics.

An alternative approach could use straightforward stochastic simulations to produce a substantial number of traces, cluster them and then classify the centroids of the clusters according to some external criteria. This is a perfectly sound method, but it requires more setup and, in general, lacks the accuracy and elegance of formal verification. It would also require an automated centroid inspection algorithm, most likely reducible to LTL checking anyway.

4 Invitation

Everyone is invited to undertake this project; we would like to do it ourselves, but are extremely busy people. We can offer guidance, generic Haskell code to produce PRISM modules, and scripts to run PRISM on a cluster. We think the project would make a very cool Master thesis, so if you like what you have just read, consider passing this document to your students.

Acknowledgements

The authors thank the Santa Fe Institute for bringing them together for the Complex Systems Summer School 2009, where this project was conceived.

References

[1] PRISM: Probabilistic Symbolic Model Checker. http://www.prismmodelchecker.org.

[2] Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking*. MIT Press, 2008.

[3] Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767), 2000.

[4] John G. Kemeny and J. Laurie Snell. *Finite Markov Chains*. Springer, 1976.

[5] M. Kwiatkowska, G. Norman, and D. Parker. Stochastic model checking. In *Formal Methods for the Design of Computer, Communication and Software Systems: Performance Evaluation (SFM'07)*, volume 4486 of *LNCS*, 2007.

[6] Orkun Soyer, Marcel Salathe, and Sebastian Bonhoeffer. Signal transduction networks: Topology, response, and biochemical reactions. *J. Theor. Biol.*, 238, 2006.

[7] Andreas Wagner. *Robustness and Evolvability in Living Systems*. Princeton University Press, 2005.