POGG: Integrating qualitative and quantitative data of macromolecular networks. A probabilistic toolbox

Giraudet Étienne, Docet Éric and Bourdon Jérémie∗
LINA CNRS FRE 2729, Nantes University
2, rue de la Houssinière
44322 Nantes (France).

1 Introduction

With the breakthrough of new techniques in experimental biology, there has been an increasing amount of data that needs to be treated, classified and analyzed. Data allow to design macromolecular networks which represent the qualitative behavior of observations as accurately as possible with experiments. On the other hand, there exists several quantitative data associated to a specific biological system (such as protein concentration variations or time series).

Several softwares are dedicated to the study of qualitative networks: PathWay Logic [4], Bioambients [7] and Biocham [3]. Their principle is to develop a formal mathematical language, to express the network in this language and to check if the resulting model satisfies some properties. Others focus on simulating the quantitative behavior of a given model: Moleculizer [6] simulates stochastic networks; BioNetGen [5] simulates ODE and stochastic networks. However, few softwares combine both qualitative and quantitative information to study model behaviors.

We overcome this lack by developing POGG software. It is based on a new treatment process for studying stochastic graph models. It introduces quantitative information into an already verified qualitative network. The core of POGG lies on toll based costs theoretical results previously studied in [2]. Such a theory assumes that evolutions of important quantities of the living system are dictated by accumulations of small elementary contributions. In our probabilistic model, the global theoretical behavior of such quantities is related to asymptotic characteristics of the stochastic process. POGG software takes profits of this theoretical knowledge and builds constraints between the observed measures and (a priori unknown) impact values of each transition. The software solves these constraints by quantifying consistent values for every impacts.

Such an information of importance allows us to reason on the model which is particularly accurate for a better understanding of living system behaviors.

2 Building a Model with POGG

POGG assumes biological models as probabilized graphs and implements an user-friendly interface for creating them. It also allows to import (and export) an existent model in several formats such as SBML or DOT. Each edge of the graph is associated to an unknown probability (the probability of taking a given edge at a given step) which corresponds to the local impact of the transition on the system. POGG reduces the search space by decreasing the range of this impact.

POGG reasons on the evolution of quantities associated to the graph. This evolution consists in an accumulation of small contributions due to each transition. This is coded by a cost matrix that is automatically created for some simple requests that answer questions like “what is the probability of passing (for the first time/once/at least once) in a given state or by a given pathway?”. The user also has the possibility to define its own cost matrix $C = (c_{i,j})$. POGG deals with different types of contributions depending on the kind of biological quantities. For transition from state $i$ to state $j$ with toll cost $C_{i,j}$, a quantity $q$ is assumed additive if its value after the transition is $q + C_{i,j}$, or multiplicative if its value after the transition is $q \times C_{i,j}$. Multiplicative transitions are particularly useful to model protein concentration evolutions.

3 Studying a Model with POGG

As major contribution, POGG deals with unknown probabilities using theoretical formulas for the mean and the variance of properly defined quantities (see [2]) for building (in-)equations between experimental quantities and asymptotic characteristics of a weighted transition matrix (e.g. dominant eigenvalue and eigenvector). This matrix combines transitions probabilities and toll costs.
Due to the size of biological networks, POGG performs a local search based on Tabu meta-heuristic for determining an appropriate probabilistic model. This approach is particularly accurate when dealing with the introduction of experimental noise. Therefore, POGG provides an approximation of the solution space volume. Such estimations are then used to simulate the behavior of other quantities that belong to the system.

Figure 1 illustrates this purpose with proteins FIS and CYA evolutions in the gene regulatory network model of Escherichia Coli carbon starvation system in the stationary growth phase (Model adapted from [8]). We assume FIS concentration as a multiplicative quantity that increases by 20% for each transition pointing to gene fis and decreases by 5% for other transitions (abstraction of the natural protein degradation). Experimental variations of FIS concentration are used to predict probabilities that compose the model. It notably predict an evolution of CYA concentration that corresponds to an already known behavior (see [1]).

4 AKNOWLEDGMENTS

This work was supported by ATLANSTIC Research Project and Ouest-Genopole. We wish to thank Damien Eveillard, Theo Merle, Christine Sinoquet and Charlotte Truchet for their useful feedback.

REFERENCES