Topological Modelling of Biological Interactions for Robustness

Danny McClanahan and William Cox
Vanderbilt University

1 Goals

Notation

To establish a framework for understanding chains of biological reactions as dynamical systems.

Analysis

To define measures of similarity and robustness for dynamical systems so that a quantifiably small change in the input does not strongly affect the output.

Application

1. To create a method of finding systems which produce the desired chemical concentrations with a high degree of robustness against changes in input.
2. To find which neighborhoods of inputs to systems lead to groups of systems with similar behavior.
3. To produce a software library to find robust alternatives to known biological reactions.
4. To produce a predictor for equilibrium points for these systems.
5. To use the predictor to identify how different input values can quantifiably affect the output behavior of a system.

Verification

1. To verify that the robustness metric defined in this study increases the resilience of a system to factors such as mutation which can modify biological networks.

2 Description as Dynamical System

Biological chemical networks can be described in terms of dynamical systems with inputs. n is the number of chemicals in the system. W is an $n \times n$ matrix

of functions which map time and concentrations of two chemicals to the amount which the concentration of the first chemical is affected. W models internal influences, such as inhibition or induction of one chemical on another (or itself). D is an n-tuple of functions which map a time t to some real number. D models external influences, such as input of a chemical and the natural degradation of a chemical over time.

Let $\mathbb{R}^+ = \{r \in \mathbb{R} \mid r \geq 0\}$. Formally, we can describe a system $j = \langle s_0, D, W \rangle$ such that $s_0 \in \mathbb{R}^n, D_i \in \mathbb{R}^{\mathbb{R}^+}, W_{ij} \in \mathbb{R}^{\mathbb{R}^+ \times \mathbb{R}}$. $s(t) \in \mathbb{R}^n$ is a state vector of chemical concentrations at time t. Let $s_i(t)$ be the i^{th} component of state vector x at time t. Y refers to the set of all such systems y. Then, $Y = \mathbb{R}^n \times (\mathbb{R}^{\mathbb{R}^+})^n \times (\mathbb{R}^{\mathbb{R}^+ \times \mathbb{R}})^n$.

$$y = \langle s_0, D, W \rangle \tag{1}$$

$$\frac{ds_i}{dt}(t) = D_i(t) + \sum_{j=1}^{n} W_{ij}(t, s_j(t)) \tag{2}$$

Note that the system is described in terms of s_0 and parameters to the differential equation $\frac{ds}{dt}$, not by a single function $s(t)$, its state vector at time t. This is because while explicit, closed-form solutions to some dynamical systems (strictly linear ones) do exist, this is not true in the general case, so obtaining a single formula for $s(t)$ is impossible. Modelling of the dynamics of these systems is possible by discretization, but there is some error between the simulated version of the system and the true $s(t)$, which can become extremely significant in some cases, particularly where the dynamics become chaotic. The effect of this discrepancy is discussed in **(where?)** when applying this model to real systems.

3 Difference Between Dynamical Systems

We define a "difference" of dynamical systems of this type to quantify the "closeness" of one network's behavior to another. This is defined in two ways. First note that there is a difference between the parameters ("input") of a system and the system's behavior ("output").

The inputs of a system are the values of $\langle s_0, D, W \rangle$. The "distance" between these is problem-dependent. For modelling a linear system, for example, D and W can be a vector and a matrix, and distance can simply be the euclidean distance, as discussed in **(where?)**. For more complex systems, "input distance" can be mapped to a normal distribution, as discussed in **(where?)**.

The "output distance" of a system is some way to quantify the distance between attractors of the system, or the behavior of the system over time. For systems with a defined convergence point, the distance between the two can be reduced to just the distance between the convergence points, as discussed in **(where?)**. For more complex periodic or aperiodic systems, other methods are usable as well, as discussed in **(where?)**.

3.1 Equilibrium System

We only concern ourselves with systems "spinning" about an "equilibrium region" for now. We create a time $t_e : Y \to \mathbb{R}$ (e for "equilibrium", even though actual equilibrium is constant) to represent the time at which a system begins spinning about an equilibrium region.

Given some $\epsilon > 0$, and t_f a final time point at which we stop observing the system, we define $t_e(y)$ as follows:

$$\int_{t_e(y)}^{t_f} |s(t) - s(t_e(y))|\, dt < \epsilon \tag{3}$$

EXPLAIN THAT t_f IS REQUIRED SO THAT THE INTEGRAL CONVERGES
ADD IMAGE EXPLAINING VISUALLY HOW WE DEFINE SOME OF THESE TERMS

3.1.1 Definition of Difference

For two systems $y_1, y_2 \in Y$, we define the distance M (M for "metric", even though this isn't a metric (I think)). For equilibrium systems, we denote this M_{eq}. s_e is the average of all points past t_e, and represents the equilibrium point of the system.

$$s_e : Y \to \mathbb{R}^n \tag{4}$$

$$s_{e_i}(y) = \frac{\int_{t_e(y)}^{t_f} s_i(t)\, dt}{t_f - t_e(y)} \,\forall\, i \tag{5}$$

$$M : Y \times Y \to \mathbb{R}^+ \tag{6}$$

$$M_{eq}(y_1, y_2) = |s_e(y_1) - s_e(y_2)| \tag{7}$$

3.2 Periodic System

DO AN EXAMPLE OF A PERIODIC SYSTEM AS WELL ALLOW SHIFTING AND STRETCHING

4 Robustness of a Dynamical System

"Robustness" is meant to quantify the tendency of the system to remain similar (the "similarity" defined in 3) given small changes in input. Robustness R is described as the inverse of a fragility F. There is a simple formulation F_s used to motivate a more general formulation F_g.

$$F : Y \to \mathbb{R}^+ \tag{8}$$
$$R : Y \to \mathbb{R}^+ \tag{9}$$
$$R(y) = F(y)^{-1} \tag{10}$$

4.1 Simple Robustness

The simple formulation quantifies the condition that all "nearby" systems have "similar" behavior. This is extremely similar to the definition of continuity of a function at a point in a topological space. Nearby systems to a state are those within the ball centered at that state, for a given radius $\epsilon > 0$, using the euclidean metric for *initial state vectors*; this metric is denoted as m, separate from the M defined in (7). s_{i_0} refers to the initial state vector s_0 for the system y_i. p The below definition of fragility describes the radius δ (defined by the difference M in (7)) required to contain all systems within radius ϵ (defined by the metric m described above).

$$m(y_1, y_2) = |s_{1_0} - s_{2_0}| \tag{11}$$
$$F_s(y) = \underset{\delta > 0}{\operatorname{argmin}}\, M(y, y') < \delta \,\forall\, y' \in B_m(s, \epsilon) \tag{12}$$

4.2 General Robustness

The general formulation of fragility of a system is defined in terms of three functions f, g, and h. f "weights" each possible pair of initial states, g "weights" the difference between each nearby system's behavior (and therefore typically makes use of the distance M from (7)), and h combines the two measures.

$$f : Y \times Y \to \mathbb{R} \tag{13}$$
$$g : Y \times Y \to \mathbb{R} \tag{14}$$
$$h : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \tag{15}$$
$$F_g(y) = \int_{y' \in Y} h(f(y, y'), g(y, y')) \tag{16}$$

As an example of the utility of F_g, it can be used to model the simple formulation F_s, given some $\epsilon > 0$.

$$M_{max}(y) = \max_{\sigma \; s.t. \; m(s_0,\sigma)<\epsilon} M(y,\sigma) \tag{17}$$

$$f(y,y') = 1 \tag{18}$$

$$g(y,y') = \begin{cases} 1 & M(y,y') < M_{max}(y) \\ 0 & else \end{cases} \tag{19}$$

$$h(a,b) = a \cdot b \tag{20}$$

These general formulations F_g and R_g become extremely useful when characterizing complex or stochastic behavior, as seen in 5.

5 Finding Maximally Robust Similar Systems

Given (1), (2), (7), and (16), we can attempt to solve the problem posed in the introduction. We wish to find a system y^* which is similar to some system y and which is maximally robust for some expected values of inputs.

Given $\epsilon > 0$. Let $y' = \langle s_0', D, W \rangle$ for some $s_0' \in \mathbb{R}^n$. This means y' has the same parameters D and W as a network $y \in Y$, but a different initial state.

THIS ASSUMES ONLY EQUILIBRIUM WITH $M = |\cdot|$ FIX THAT

$$Y_{similar} = \{y' \in Y \mid M(y,y') < \epsilon\} \tag{21}$$

$$y^* = \operatorname*{argmin}_{y' \in Y} F(y') \tag{22}$$

5.1 Modelling Stochastic Distributions Using a Weight Function

In 4.2, we described how we use the function f to "weight" certain pairings of systems, in a sort of distance-like metric. This can be used to model stochastic distributions of parameters.

Recall that a system $y \in Y$ is described by a tuple $\langle s_0, D, W \rangle$. We shall assume that each parameter arises from a normal distribution with mean μ and standard deviation σ. This produces vectors $\mu_y \in \mathbb{R}^{2n+n^2}, \sigma_y \in \mathbb{R}^{2n+n^2}$ ($2n + n^2$ is the number of all parameters in a system y, as described in 2). We can then make an f which weights other systems y' according to how far they deviate from their mean in μ_y.

6 Simulation Methods

TALK ABOUT HOW WHAT WE'RE DOING HERE IS BASICALLY A BIG MONTE CARLO SIMULATION WHICH HAS LOTS OF PRECEDENT IN LOTS OF PLACES