

Topological Modelling of Biological Interactions for Robustness

1 Goals

1. To define a measure of robustness of network topologies so that small changes in the input do not affect the output.
2. To create software which allow searching for networks which produce the desired chemical concentrations with a high degree of robustness against changes in input.

2 Theory

A topology is a collection of n (currently $n \leq 3$) chemicals which can affect (increase/decrease the concentration of) each other and themselves. Complex biological interactions can often be modelled with network topologies. These are typically described with differential equations. Over time, relative concentrations of these chemicals will either reach equilibrium, form a periodic or chaotic pattern, or become unstable.

The **robustness** of a topology is defined (qualitatively) to be the amount it resists changes in its output given changes in its input. Inputs, outputs, and a quantitative description of robustness are given in the sections below.

2.1 Topology

A biological interaction topology is defined to be a weighted directed graph (with loops). There might be a better word for it, but that's the one we use here. We will represent this as a 3x3 matrix. Each node corresponds to a concentration of some chemical, and each edge corresponds to the effect each chemical has on the concentration of another, through protein-protein interactions (PPIs), transcription interactions, or other biological nonsense. Interactions are represented as a real number on the interval $[-1, 1]$. Positive numbers correspond to induction, zero corresponds to no interaction, and negative numbers correspond to repression. This matrix is called W , and the interaction of chemical i on chemical j is W_{ij} . We allow loops to simulate self-interactions.

2.1.1 Topologies in Software

Topologies are generated by exhaustively filling each element of the 3x3 matrix with all permutations of real numbers on the unit interval representable using

an IEEE 754 double-precision floating-point number, with a certain granularity $G \in \mathbb{N}$. All numbers in a given topology are selected through the formula $\frac{k}{G}$, where $k \in [-G, G] \subseteq \mathbb{Z}$.

2.2 Sample

A sample of a topology is a topology combined with certain other factors. These factors currently include:

1. initial concentrations vector $C_{initial} \in \mathbb{R}^n$
2. hill coefficient H (not yet implemented) [1]
3. dissociation constant S (not yet implemented) [1]
4. degradation factor vector $D \in \mathbb{R}^n$

All these factors are normalized and drawn from the unit interval with granularity G as described in 2.1. They all affect the output concentrations as described in (1). They can be described as a vector, as can the output concentrations. The use of sample input as a vector is described in 3. The degradation factor can also model a constant input of a chemical into the system by just making that entry D_i of D negative.

3 Methods

We define robustness for both a topology y and for a particular sample α of y .

3.1 Input Factors Versus Topology Weights

The use of “input” and “output” factors as a vector has been referred to, but not explained. Topology weights 2.1 and input factors 2.2 can both be viewed as vectors. They are separated because topology weights are fixed by biology and chemistry, while input factors can be influenced by, say, synthetic biologists more easily. The idea is that for a desired set of chemicals, the synthetic biologist can modify the values of input factors to get a desired level of robustness, while the desired set of chemicals chosen determine the topology weights.

3.2 Samples as Vectors

The simulation as described in 3.4 goes from a vector of initial concentrations $C_{initial}$ to a vector of output concentrations C_T , for each sample of a topology. The “input vector” is then the concatenation of $C_{initial}$ and the degradation factor vector D (and anything else mentioned in 2.2). We denote the length of I as p .

3.3 Notation

Let the set of all topologies be denoted Y . For a topology $y \in Y$, let A be the set of all samples of y . For a sample $\alpha \in A$, let I_α be the input vector for the sample α , and O_α be the output vector. Recall that $O_\alpha = C_T$ as described in 3.4. Let α_i be the i^{th} sample of A .

3.4 Simulation

For a continuous formulation of C_T , we describe as a differential equation:

$$\frac{dC_i}{dt} = -D_i + \sum_{j=1}^n W_{ij} \quad (1)$$

In a discrete simulation context, the initial concentrations of n nodes can be viewed as a vector $C \in \mathbb{R}^n$. Simulation occurs by “stepping” the concentration vector. C_t is a time-varying vector, changing at each time t , depending upon the previous value C_{t-1} . Stepping refers to applying the formula in (1) discretely to each concentration $C_{t_i} \in C$. The initial concentrations $C_{initial}$ are used at $t = 0$, so $C_0 = C_{initial}$, and the next step can be described:

$$C_{(t+1)_i} = C_{t_i} (1 - D_i + \sum_{j=1}^n W_{ij}) \quad (2)$$

This simulates the differential equation governing time-varying concentrations. The “output” of a sample is the vector $C_T \in \mathbb{R}^n$, where T is the number of timesteps in the simulation. T is constant over all trials. t is an arbitrary timestep.

3.5 Normalization

Normalizing each number produces all valid topologies and samples, because larger ones can be viewed as scaled-up (dilated, multiplied, whatever you want to call it) versions of smaller ones. Increasing the granularity G allows for more larger samples to be viewed as scaled-up version of smaller topologies with all input parameters within the unit interval. This is allowable because all equations used are strictly linear.

However, when taking into account Hill coefficients, this linearity may vanish. We’re not sure whether this applies because we haven’t implemented Hill coefficients yet.

3.6 Robustness of a Topology

Robustness of a topology is defined in a coarse-grained fashion. Let $|A|$ be the cardinality of A , and $r : Y \rightarrow \mathbb{R}$ the robustness function of a topology y . Then, O_j is a list of all the j^{th} elements of $O_\alpha \forall \alpha \in A$, and σ_j is the standard

deviation of that list. This assumes a normal distribution, which is allowed for some reason I don't remember right now.

The robustness $r : Y \rightarrow \mathbb{R}$ of the topology y is defined to be the euclidean norm of O_σ . Recall that A and O_σ are only defined in reference to a particular topology y producing samples $\alpha \in A_y$.

$$O_j = \{O_{\alpha_{1j}}, O_{\alpha_{2j}}, \dots, O_{\alpha_{|A|j}}\} \forall \alpha_i \in A_y \quad (3)$$

$$\sigma_j = \sigma(O_j) \quad (4)$$

$$O_{\sigma_j} = \sigma_j \quad (5)$$

$$O_\sigma = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_{|A|} \end{bmatrix} \quad (6)$$

$$r(y) = |O_\sigma| \quad (7)$$

We are not certain whether this definition makes sense in a continuous context.

3.7 Robustness of a Sample

Robustness of a sample $s : A_y \rightarrow \mathbb{R}$ is first defined naïvely, then in a more nuanced and complete way. The intuition for robustness is that for small variations in input parameters of a sample α , the variation in the output parameters is also small. This sounds similar to the definition of continuity at a point for topological spaces, although it is actually the reverse of that definition ($\forall \epsilon \exists \delta$, not vice versa).

3.7.1 Naïve Robustness of a Sample

Start by defining metrics d_I and d_O on samples $\alpha, \beta \in A$.

$$d_I, d_O : A \times A \rightarrow \mathbb{R} \quad (8)$$

$$d_I(\alpha, \beta) = |I_\alpha - I_\beta| \quad (9)$$

$$d_O(\alpha, \beta) = |O_\alpha - O_\beta| \quad (10)$$

d_I is conceptually the euclidean distance between the input vectors for samples $\alpha, \beta \in A_y$ for some $y \in Y$. d_O is defined similarly.

Naïve robustness of a sample is defined to be: for a given (constant) $\epsilon > 0$, the $\delta > 0$ such that $\forall \beta \text{ s.t. } \beta \in B_{(d_I)}(\alpha, \epsilon), \beta \in B_{(d_O)}(\alpha, \delta)$. $B_f(p, r)$ is defined to be the neighborhood about point p with radius r using the metric defined by f . This is a relative measure, so a “robust” sample is one with a δ below some arbitrary threshold value. To restate, for some $\epsilon > 0$:

$$s(\alpha) = \operatorname{argmin}_{\delta > 0} \forall \beta \in B_{(d_I)}(\alpha, \epsilon), \beta \in B_{(d_O)}(\alpha, \delta) \quad (11)$$

Smaller is better.

3.7.2 Decaying Robustness of a Sample

The previous definition of robustness does not take into account samples $\beta \in A$ that are “far” away from some sample α (anything with $d_I \geq \epsilon$). The motivation for a more nuanced definition is that even with the assumption that most input parameters (initial concentrations, degradation, etc.) will remain relatively close in a real-world context to the vector I_α (meaning that most input vectors will be normally distributed with mean I_α), it would be preferable to create a measure which is influenced by those “far” points, just less so than the more likely “near” points.

Let $f, g : A \times A \rightarrow \mathbb{R}$. Conceptually, f and g are functions which assign a “weight” to another sample by performing some function on the difference of two vectors. f acts upon input vectors I , and g acts upon output vectors O . Both f and g can be thought of as compositions of a function f' or g' with d_I or d_O , such that:

$$f', g' : \mathbb{R} \rightarrow \mathbb{R} \quad (12)$$

$$f = f' \circ d_I \quad (13)$$

$$g = g' \circ d_O \quad (14)$$

It is possible to choose f, g which are not decomposable into $f' \circ d_I$ or $g' \circ d_O$, but we have not investigated these classes of f and g .

f' and g' should be functions which weight “closer” vectors greater than “further” ones. For example, for some $\epsilon, \delta > 0$:

$$f'(r) = \begin{cases} 1 & r < \epsilon \\ 0 & r \geq \epsilon \end{cases} \quad (15)$$

$$g'(r) = \begin{cases} 1 & r < \delta \\ 0 & r \geq \delta \end{cases} \quad (16)$$

The above is equivalent to the naïve definition of robustness of a sample. Other choices of f' and g' could include asymptotically decaying functions such as exponential functions of the form $f'(r) = Ae^{-br}$.

We then define robustness of a sample $\alpha \in A$ as follows.

$$s(\alpha) = \int_0^1 \cdots \int_0^1 ((f(\alpha, \beta) \cdot g(\alpha, \beta))^{-1}) dI_{\beta_1} \cdots dI_{\beta_p} \quad (17)$$

Recall that p is the number of elements in the vector I .

To define this over a discretization of all ranges of the input vectors I_β :

$$s(\alpha) = \sum_{\beta \in A} (f(\alpha, \beta) \cdot g(\alpha, \beta))^{-1} \quad (18)$$

As in the naïve formula for robustness of a sample, this produces an arbitrary value, which must be judged relative to the value produced by other samples. Unlike the naïve formulation, bigger is better.

Consider flipping the exponent to +1?

4 Aside on Variance Over Time

Differential equations are used here to model time-varying concentrations of different chemicals. However, the time variance is not discussed here. This is because dynamical systems such as this either reach equilibrium or become unstable, and unstable systems will exhibit a very high degree of variance. Therefore, when searching for robust networks which satisfy certain conditions, these systems will not be found, since they are not robust.

5 Analytical Results

Use differential equations and dynamical systems theory?

6 Simulation Results

Find a good granularity G and maximum timestep T and we're golden.

References

- [1] iGEM UNIPV-PAVIA. Mathematical modeling page, 2008. [Online; accessed 3-March-2016].