Physical Modelling of Complex Systems: Assignment 4

Thibeau Wouters

April 5, 2021

Contents

1 Motifs in the E. Coli Gene Regulatory Network

The regulatory network of the bacteria E. Coli is given to us in a `txt` file, where a number of different proteins, represented as nodes, are in interaction with each other, represented as edges. The file `coliInterFullNames.txt` gives a dictionary linking the number value to the name of the protein. From this, we see that there are 424 different proteins in the E. Coli bacteria which are relevant for our problem. Hence the number of nodes of the network is $N = 424$. The number of edges is given by the number of rows in the file `coliInterFullVec.txt`, and we find that $E = 577$ for our problem.

1.1 Selfregulating proteins

We are first of all interested in the number of selfregulating proteins N_{self} in the E. Coli network. This number can be determined via a Python script, which checks whenever the

number in the x and y column match with each other, indicating that this row represents an interaction of a certain protein with itself. There are 58 selfregulating proteins: 14 are self activators, 41 are self repressors and for 3 selfregulating proteins, the type of reaction is unknown.

We can analytically calculate the average number of selfregulating nodes in a random network, which we'll denote as $\langle N_{\text{self}} \rangle$. The following derivation is based on chapter 3 from Uri Alon's book [1].

In a random network, we choose a pair of nodes (X, Y) at random, of which one is the origin, and the other is the destination. There is a probability of $p_{\text{self}} = 1/N$ to form a self-edge (i.e. a pair of nodes with $X = Y$). Since there are E edges, the probability of having precisely k self-edges is therefore given approximately by a binomial distribution P with probability of success p_{self}, such that

$$P(k) = \binom{E}{k} p_{\text{self}}^k (1 - p_{\text{self}})^{E-k} \,. \tag{1.1}$$

The reason that it is *approximately* binomial is due to the fact that we ignored above that we cannot make two self-edges using the same node and hence p_{self} depends on the number of self-edges already formed. For sparse networks (defined later on), of which the E. Coli network is an example, p_{self} does not change much and the above probability distribution is a good approximation. The mean of the above distribution is

$$\langle N_{\text{self}} \rangle \approx E p_{\text{self}} \sim E/N \,. \tag{1.2}$$

For the parameters of the E. Coli network, this gives

$$\langle N_{\text{self}} \rangle \approx 1.3608 \,. \tag{1.3}$$

We can also compare this by creating 2000 random networks (see later on) and compute the average number of selfregulating proteins. On one of those runs, we found a value 1.3550, which is quite close to the analytically determined value.

We see that the E. Coli network has much more selfregulating proteins than expected for a random network, and hence we conclude that selfregulation is a network motif.

1.2 Feed forward loops

Another interesting regulation pattern is the feed forward loop (FFL). Note that we only consider the loops where all three proteins are different from each other as FFL's. To find the feed forward loops, we will create a dictionary in which every protein is represented as a key. The value corresponding to the key is a list of all the proteins that are activated by the 'key' protein. Using this, we can count the number of FFL's using the following logic. Let

X, Y and Z be three proteins. Denote by P_X the set of all proteins activated by protein X. Then we count the number of FFL's N_{FFL} as follows:

$$\forall X : \forall Y \in P_X \setminus \{X\} : \forall Z \in P_Y \setminus \{X, Y\} : \text{if } Z \in P_X \Rightarrow N_{\text{FFL}} += 1 \tag{1.4}$$

Such a function is implemented in Python. In the E. Coli file, we then find 23 CHECK THIS FFL's.

We can analytically compute the expected number of FFL's in a random network. The following derivation is based on chapter 4 of Uri Alon's book [1].

We form random networks by inserting E edges between the N nodes. The total number of edges possible is $N(N-1) + N = N^2$, where the first term represents edges between different nodes, while the second term represents self-edges. Hence the probability for a certain edge to occur is $p = E/N^2$. The E. Coli network is an example of a *sparse* network, for which it holds that $p \ll 1$. Indeed, for the E. Coli network, we have $p = 0.0032$. Let G be the subgraph representing the FFL. This subgraph has $n = 3$ nodes and $g = 3$ edges. To generate such a subgraph G in a larger graph, we need to choose n different nodes and place g edges between them. Finally, we have to take into account the probability p_{act} that each of the edges in the FFL is an activator. In the E. Coli network, about $p_{\text{act}} = 58.06\%$ of the interactions are activators. For the FFL, we then find

$$\langle N_{\text{FFL}} \rangle = N(N-1)(N-2) p^g p_{\text{act}}^g \,. \tag{1.5}$$

For the parameters of the E. Coli network, we then find

$$\langle N_{\text{FFL}} \rangle = 0.4897 \,. \tag{1.6}$$

We will now compare this to the number of FFL's we would find in a random network. For this, a Python function is written which creates a random network with the same number of nodes and edges as the E. Coli file. A number of features from the E. Coli file are taken into account. First of all, we have to avoid that the function would give twice an interaction between two nodes in the same 'direction'. Second, we would like to create a random network of which the structure of the interactions is similar to the E. Coli network. For this, we first determine the relative amounts of activators, repressors, and unknown interactions in the E. Coli file. We find 58.06% of the interactions are activators, 36.92 % are repressors, and 5.03 % are unknown. When building our random network, we will use these numbers as probabilities to choose a certain type of interaction by generating a random number between 0 and 1. If one wants to create totally random networks, there is the option to give an equal probability for each interaction type.

The Python function then works schematically as follows: choose a pair of numbers (X, Y), both between 1 and the desired number of nodes N. If the pair (X, Y) was already generated, discard and repeat this step (note that order matters, since this gives the 'direction' of interaction, namely from X to Y). If the pair (X, Y) was not yet generated, choose

an interaction based on the given probabilities for them to occur. Repeat until the desired number of edges is reached.

Creating 2000 random networks, we find that the average number of FFL's is 0.4880. This is quite close to the analytically determined value. We see that the number of FFL's in the E. Coli network is much higher than the average for random networks, and hence we can conclude that FFL is a network motif.

2 Positively autoregulated protein

The previous problem showed in an explicit example that autoregulation is a network motif. We consider here the case of a positively autoregulated protein whose production is governed by the following differential equation

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \frac{\beta X^n}{X^n + K^n} - \alpha X \,. \tag{2.1}$$

The first term is a Hill function with Hill coefficient n. We first non-dimensionalize the above differential equation by rescaling X and t appropriately. Assume that X is a concentration. Then α has the dimensions of an inverse time, and K the dimensions of a concentration. Define

$$\tau \equiv \alpha t, \quad U \equiv \frac{X}{K} \,. \tag{2.2}$$

Then the above differential equation becomes

$$\frac{\mathrm{d}U}{\mathrm{d}\tau} = \frac{\beta}{\alpha K} \frac{U^n}{U^n + 1} - U \,, \tag{2.3}$$

which is a differential equation depending on dimensionless coefficients. The dynamics is governed by a single parameter, namely $\gamma \equiv \beta/(\alpha K)$.

Let us first restrict to the case $n = 1$ in the differential equation (2.1), and look for fixed points of the equation. Thus we want to solve

$$\beta X - \alpha(X + K)X = 0 \,, \tag{2.4}$$

of which the solutions are $X = 0$ and

$$X^* = \frac{\beta - \alpha K}{\alpha} \,. \tag{2.5}$$

We see that for $\beta > \alpha K$, the fixed point is positive. Defining the right hand side of (2.1) as $F_1(X)$ for $n = 1$, we find that the derivative is

$$F_1'(X) = \frac{\beta K}{(X + K)^2} - \alpha \,, \tag{2.6}$$

and hence, we can find

$$F_1'(X^*) = \alpha \left(\frac{\alpha K}{\beta} - 1 \right) . \tag{2.7}$$

For the values of the parameters in which the fixed point X^* is greater than zero, which is $\beta > \alpha K$, this is a stable fixed point since the above derivative is negative. Note that we will always assume $\alpha > 0$, since the second term in equation (2.1) represents the degradation of the protein.

We now solve the above differential equation numerically. We set the parameters to $\alpha = 1$, $\beta = 1$ and $K = 1/4$ and take as initial condition $X(0) = 10^{-3}$. Note that this satisfies $\beta > \alpha K$, and hence there is a stable fixed point X^* which is strictly positive. Using equation (2.5), we find that $X^* = 0.75$. We will compare the solution for the above differential equation, representing an autoregulated protein, with the case of a non-regulated protein, for which the differential equation was already seen before:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \beta - \alpha X \, . \tag{2.8}$$

We would like to recall (as is readily seen from the above differential equation), that the non-regulated protein differential equation has a fixed point for $X = \beta/\alpha$, and that this is always a stable fixed point. The analytic solution to the above differential equation was computed in the lectures, and we found

$$X(t) = \frac{\beta}{\alpha} \left(1 - e^{-\alpha t} \right) . \tag{2.9}$$

Below, we take the same parameters for the non-regulated protein as for the auto-regulated protein, and hence the non-regulated equation has a fixed point at $X^* = 1$. The numerically obtained solutions for both differential equations are shown in Figure 2.1 below. As expected, both solutions tend towards their fixed point.

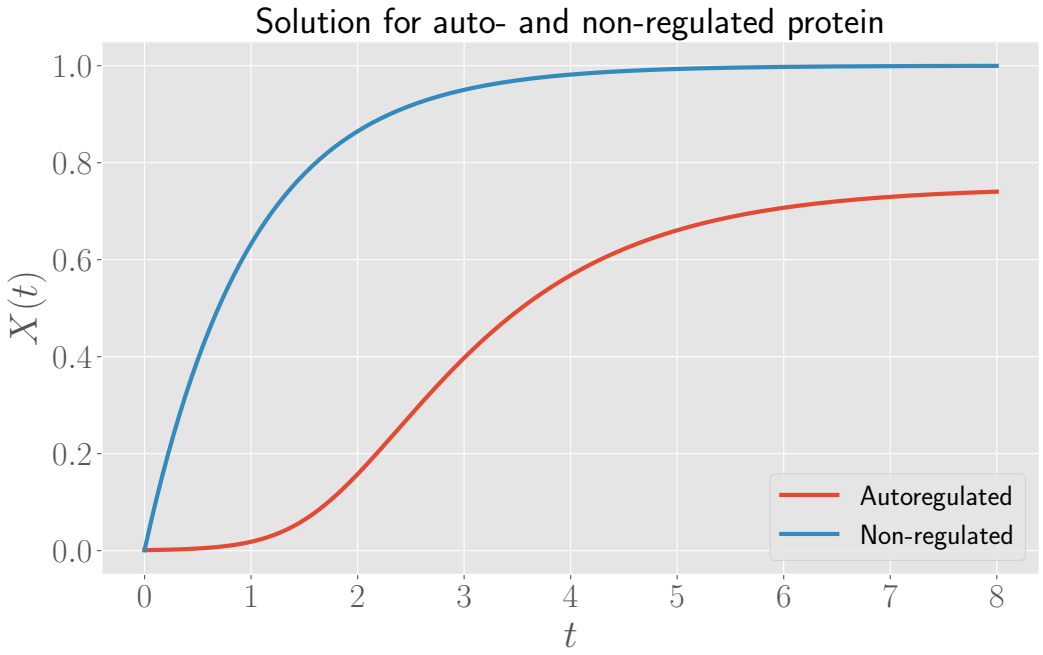

Figure 2.1: Numerical solutions for the autoregulated and non-regulated protein, with differential equations shown in (2.1) and (2.8), respectively, for the values $\alpha = \beta = 1$, $X(0) = 10^{-3}$ and $K = 0.25$.

We will now determine the response time, which is the time for the system to reach half of the stationary value concentration, for the two cases and compare them with each other. The response time for the non-regulated system can be computed using the analytic solution given above, and for the parameters currently considered, we find

$$T_R^{\text{non}} = \frac{\ln 2}{\alpha} = 0.6931\,. \tag{2.10}$$

For the autoregulated system, we will use the graph in Figure 2.1 to estimate this value (recall that the stationary concentration is at $X^* = 0.75$ for the autoregulated protein). We find

$$T_R^{\text{auto}} \approx 2.9\,. \tag{2.11}$$

Hence the response time for the autoregulated protein is much longer than the response time for the non-regulated protein.

Now let us consider the case $n = 2$ in equation (2.1). We look for the fixed points of the differential equation and hence for solutions of the equation

$$\frac{\beta X^2}{X^2 + K^2} - \alpha X = 0\,. \tag{2.12}$$

A first solution is $X = 0$. Other solutions are solutions to the differential equation

$$\alpha X^2 - \beta X + \alpha K^2 = 0\,. \tag{2.13}$$

The solutions are

$$X_{\pm}^{*} = \frac{\beta \pm \sqrt{\beta^2 - 4\alpha^2 K^2}}{2\alpha} \,. \tag{2.14}$$

However, these fixed points only appear if $\beta > 2\alpha K$, since otherwise the square root is not a real number. For this range of the parameters, there is bistability. To prove this claim, let us consider the stability of the fixed points. Define the right hand side of (2.1) to be $F_2(X)$ for $n = 2$. The stability for various values of the parameters is best studied via an applet such as Desmos. In this Desmos applet, one can plot the two individual terms in (2.1) or both terms simultaneously. The applet also prints 'bistability' or 'no bistability', depending on the specific values of the parameters. We see that whenever there are three fixed points (i.e. for $\beta > 2\alpha K$, and hence whenever the applet prints 'bistability') the origin and the fixed point X_{+}^{*} are stable, while the fixed point X_{-}^{*} is unstable. Since $\beta > \alpha K$, the fixed point X^{*} is also stable. This is indeed the behaviour we expect from a system with bistability.

3 Genetic toggle switch

We consider two transcription factors mutually repressing each other. This system can become bistable and thus can be considered as a genetic switch. To understand qualitatively bistability, we consider the case in which the concentration of X is high such that the production of Y is repressed and Y has a low concentration. If this happens, Y cannot repress X, which then maintains its high concentration. The other possibility is a symmetric situation in which the role of X and Y is interchanged. Bistability occurs only for some specific choice of model and parameters. To understand this, the model we will consider is

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \frac{\beta K^n}{Y^n + K^n} - \alpha X \,. \tag{3.1}$$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \frac{\beta K^n}{X^n + K^n} - \alpha Y \,. \tag{3.2}$$

We first perform a non-dimensionalization of the above differential equations. Assume that X and Y carry the dimensions of concentration. From the above equations, we see that α then is an inverse time, K has the dimension of concentration, and hence β carries dimensions concentration over time. So define

$$\tau \equiv \alpha t, \quad U \equiv \frac{X}{K}, \quad V \equiv \frac{Y}{K} \,. \tag{3.3}$$

Then the above differential equations become

$$\frac{\mathrm{d}U}{\mathrm{d}\tau} = \frac{\beta}{\alpha K} \frac{1}{V^n + 1} - U \tag{3.4}$$

$$\frac{\mathrm{d}V}{\mathrm{d}\tau} = \frac{\beta}{\alpha K} \frac{1}{U^n + 1} - V \,, \tag{3.5}$$

and these differential equations indeed have prefactors which are dimensionless. The form of these equations is similar to the non-dimensionalized equation in the previous exercise.

We now show that if $n = 1$, there is no bistability. For this, we will show that the above differential equations have a single fixed point (X^*, Y^*) with $X^* = Y^*$. The fixed point is a solution of the system of equations

$$\begin{cases} \beta K - \alpha(Y + K)X & = 0 \\ \beta K - \alpha(X + K)Y & = 0 \end{cases} . \tag{3.6}$$

By demanding that the second term in both equations equal each other, we easily derive the equation $X = Y$. Using $X = Y$ in the first equation, we find the equation

$$\alpha Y^2 + \alpha K Y - \beta K = 0 \, . \tag{3.7}$$

This quadratic equation has two solutions, but we will only keep one of the solutions since the other one is negative, which is physically impossible. The solution is

$$X^* = Y^* = \frac{-\alpha K + \sqrt{\alpha^2 K^2 + 4\alpha\beta K}}{2\alpha} \, . \tag{3.8}$$

Hence we can conclude that for $n = 1$, there is one fixed point (X^*, Y^*) which satisfies $X^* = Y^*$.

Now we consider the case $n = 2$. The fixed points are solutions of the system of equations

$$\begin{cases} \beta K^2 - \alpha X(Y^2 + K^2) & = 0 \\ \beta K^2 - \alpha Y(X^2 + K^2) & = 0 \end{cases} . \tag{3.9}$$

Like before, we can demand that the second term in the above equations are equal to each other. This gives the equation

$$XY^2 + K^2 X = X^2 Y + K^2 Y \, , \tag{3.10}$$

and from this, we find

$$XY(Y - X) = K^2(Y - X) \, . \tag{3.11}$$

One solution is given by $X^* = Y^*$, which is a symmetric solution. If we assume that the fixed point is not symmetric, i.e. that $X^* \neq Y^*$, we find that the asymmetric fixed points satisfy $X^* Y^* = K^2$.

These considerations allow us to explicitly find the fixed points. For the symmetric fixed points, we replace one of the above equations by $X = Y$. Using the first equation above then results in

$$\alpha X^3 + \alpha K X - \beta K^2 = 0 \, . \tag{3.12}$$

The solution to this equation is obtained via the cubic formula, as given for example here, and gives

$$X^* = \sqrt[3]{\frac{\beta K^2}{2\alpha} + \sqrt{\frac{\beta^2 K^4}{4\alpha^2} + \frac{K^6}{27}}} + \sqrt[3]{\frac{\beta K^2}{2\alpha} - \sqrt{\frac{\beta^2 K^4}{4\alpha^2} + \frac{K^6}{27}}}\,, \tag{3.13}$$

which is positive for all values of the parameters. For the asymmetric fixed points, we replace one of the equations by $XY = K^2$, and hence $Y = K^2/X$. Substituting this in the first equation, we find the following quadratic equation for X

$$\alpha X^2 - \beta X + \alpha K^2 = 0\,. \tag{3.14}$$

The solutions of this equation are

$$X_\pm = \frac{\beta \pm \sqrt{\beta^2 - 4\alpha^2 K^2}}{2\alpha}\,, \tag{3.15}$$

and the relation $XY = K^2$ then gives the Y component of these fixed points. We will refer to the fixed points (X_+, Y_+), (X_-, Y_-) as the 'asymmetric fixed points', and refer to (X^*, Y^*) as the 'symmetric fixed point'. We see that these asymmetric fixed points only exist for $\beta > 2\alpha K$. For $\beta < 2\alpha K$, there is only one symmetric fixed point. This is verified via this Desmos applet.

To determine the nature of the fixed points, we have to calculate the Jacobian matrix. The Jacobian matrix is

$$J(X, Y) = \begin{pmatrix} -\alpha & -\frac{2\beta K^2 Y}{(Y^2+K^2)^2} \\ -\frac{2\beta K^2 X}{(X^2+K^2)^2} & -\alpha \end{pmatrix}\,. \tag{3.16}$$

The trace of the Jacobian matrix is -2α, and its determinant is

$$\det J(X, Y) = \alpha^2 - 4\frac{\beta^2 K^4 XY}{(X^2+K^2)^2(Y^2+K^2)^2}\,. \tag{3.17}$$

In the aforementioned Desmos applet, we can plot the determinant and trace of the Jacobian evaluated for the one or three fixed points (depending on the range of parameters) in a $(\det A, \operatorname{tr} A)$ plane, along with the curve $\operatorname{tr} A = \frac{1}{4}(\det A)^2$ like we did in the lectures. Note that for all three fixed points, the trace is always equal to -2α. We will now show that the for the asymmetric fixed points, the determinant of the Jacobian evaluated at (X_+, Y_+) and (X_-, Y_-) are identical to each other. First of all, note that for these fixed points, we have[1]

$$\begin{aligned}(X^2+K^2)^2(Y^2+K^2)^2 &= \left(X^2+K^2\right)^2\left(\frac{K^4}{X^2}+K^2\right)^2 \\ &= \frac{K^4}{X^4}\left(K^2+X^2\right)^4 \\ &= K^4\left(X+\frac{K^2}{X}\right)^4\,,\end{aligned} \tag{3.18}$$

[1]We will drop the plus and minus indices to simplify the notation a bit.

where in the first step, we used the defining relation for the asymmetric fixed points $XY = K^2$. As an intermediate step, we can show that

$$\begin{aligned} X + \frac{K^2}{X} &= \frac{\beta \pm \sqrt{\beta^2 - 4\alpha^2 K^2}}{2\alpha} + \frac{2\alpha K^2}{\beta \pm \sqrt{\beta^2 - 4\alpha^2 K^2}} \\ &= \frac{(\beta \pm \sqrt{\beta^2 - 4\alpha^2 K^2})^2 + 4\alpha^2 K^2}{2\alpha \left(\beta \pm \sqrt{\beta^2 - 4\alpha^2 K^2}\right)} \\ &= \frac{2\beta^2 \pm 2\beta\sqrt{\beta^2 - 4\alpha^2 K^2}}{2\alpha(\beta \pm \sqrt{\beta^2 - 4\alpha^2 K^2}} = \frac{\beta}{\alpha}\,, \end{aligned} \tag{3.19}$$

and hence we can conclude that

$$(X^2 + K^2)^2 (Y^2 + K^2)^2 = K^4 \left(\frac{\beta}{\alpha}\right)^4 . \tag{3.20}$$

Using our earlier expression for the determinant, we find

$$\det J(X_\pm, Y_\pm) = \alpha^2 - 4\frac{\alpha^4 K^2}{\beta^2}\,, \tag{3.21}$$

which is indeed the same for both asymmetric fixed points.

Playing around with the parameters in the Desmos applet, we see that the asymmetric fixed points are always stable, while the symmetric fixed point is a saddle node for $\beta > 2\alpha K$ and stable for $\beta < 2\alpha K$.

References

[1] Uri Alon. *An introduction to systems biology: design principles of biological circuits*. CRC press, 2019.