

1.

a.

Set of Variables:

G^* : expression level of the gene of interest.

G_1, G_2, \dots, G_m : expression levels of upstream genes.

m : the number of conditions.

Given that a linear function can relate activities of the gene G^* and the set of upstream genes G_1, G_2, \dots, G_m , we can model the expression level of G^* for a condition i as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in} + \epsilon_i$$

where y_i is observed gene expression level of G^* in condition i .

x_{ij} is expression level of gene G_j in condition i .

β_0 is the intercept

$\beta_1, \beta_2, \dots, \beta_n$ are coefficients for the corresponding upstream gene

ϵ_i is the error term and $\epsilon_i \sim N(0, \sigma^2)$

Since $\epsilon_i \sim N(0, \sigma^2)$

$$\epsilon_i = y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in})$$

$$\Rightarrow f(x_i | \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{(y_i - (\beta_0 + \sum_{j=1}^n \beta_j x_{ij}))^2}{2\sigma^2}\right)} \text{ for a given condition } i.$$

The Likelihood for all observations:

$$\mathcal{L} = \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{(y_i - (\beta_0 + \sum_{j=1}^n \beta_j x_{ij}))^2}{2\sigma^2}\right)}$$

Log-likelihood for all observations:

$$\log(L) = m \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2} \sum_{i=1}^m \left(y_i - \left(\beta_0 + \sum_{j=1}^n \beta_j x_{ij} \right) \right)^2$$

$$\Rightarrow \log(L) = -\frac{m}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m \left(y_i - \left(\beta_0 + \sum_{j=1}^n \beta_j x_{ij} \right) \right)^2$$

To find the values of $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ that maximize the log-likelihood function, the objective function then becomes:

$$\Rightarrow \max(\log(L))$$

$$\Rightarrow \max -\frac{m}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m \left(y_i - \left(\beta_0 + \sum_{j=1}^n \beta_j x_{ij} \right) \right)^2$$

⑥ Set of Variables:

$G^*(t)$: expression level of gene of interest G^* at time point t .

$G_1(t), G_2(t), \dots, G_m(t)$: expression levels of upstream genes

G_1, G_2, \dots, G_m at time point t .

$\gamma_0, \gamma_1, \gamma_2, \dots, \gamma_n$: rate constants of the ODEs

Given that ODEs describe the dynamics of the expression level of G^* which is influenced by the expression levels of upstream genes G_1, G_2, \dots, G_m , we can develop a model as follows:

$$\frac{dG^*}{dt} = r_0 + r_1 G_1 + r_2 G_2 + \dots + r_n G_n$$

$$\frac{dG_1}{dt} = f_1(G^*, G_1, G_2, \dots, G_n, r_1, r_2, \dots, r_n)$$

\vdots

$$\frac{dG_n}{dt} = f_n(G^*, G_1, G_2, \dots, G_n, r_1, r_2, \dots, r_n)$$

where f is a function that describes how the rate of change in expression of G^* is influenced by the expressions of the upstream genes and the rate constants.

Assuming a normally distributed error term ($\epsilon_i \sim N(0, \sigma^2)$) again, the likelihood function of all observations becomes:

$$\mathcal{L} = \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{(y_i - G^*(t_i))^2}{2\sigma^2}\right)}$$

where y_i is the observed expression level of G^* at time point i

$$\log(\mathcal{L}) = -\frac{m}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m (y_i - G^*(t_i))^2$$

Objective is to maximize the $\log(\mathcal{L})$

$$\max \left(-\frac{m}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m (y_i - G^*(t_i))^2 \right)$$

2.

$$\textcircled{a} \quad \frac{\partial L}{\partial w_1} = 2w_1 + 2$$

$$\frac{\partial L}{\partial w_2} = 4w_2 - 4$$

$$\frac{\partial L}{\partial w_3} = 2w_3$$

$$\frac{\partial L}{\partial w_4} = -2w_3 + 2w_4$$

$$\nabla L(w) = \begin{bmatrix} \partial L / \partial w_1 \\ \partial L / \partial w_2 \\ \partial L / \partial w_3 \\ \partial L / \partial w_4 \end{bmatrix} = \begin{bmatrix} 2w_1 + 2 \\ 4w_2 - 4 \\ 2w_3 - 2w_4 \\ -2w_3 + 2w_4 \end{bmatrix}$$

$$\textcircled{b} \quad \nabla L(w) = 0$$

$$2w_1 + 2 = 0 \Rightarrow w_1 = -1$$

$$4w_2 - 4 = 0 \Rightarrow w_2 = 1$$

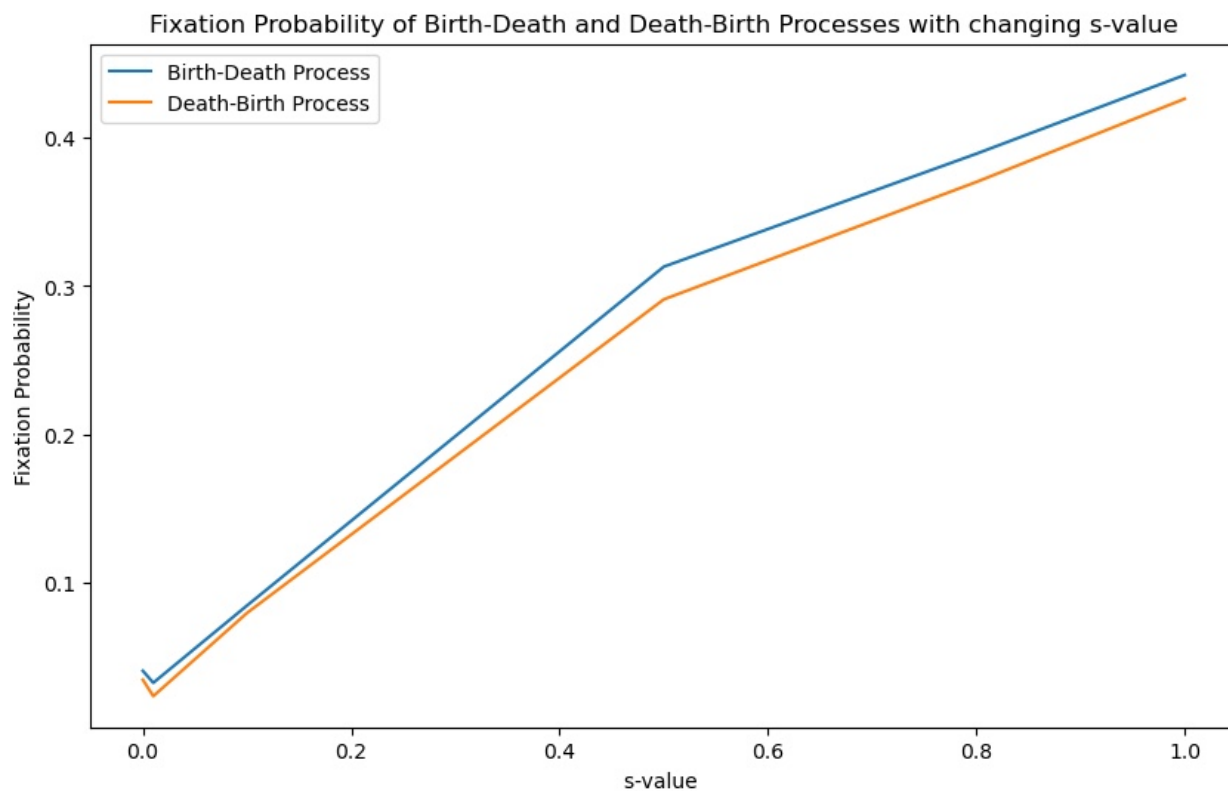
$$2w_3 - 2w_4 = 0 \Rightarrow w_3 = w_4$$

$$-2w_3 + 2w_4 = 0 \Rightarrow w_3 = w_4$$

$$\textcircled{c} \quad \nabla L(w) = 0$$

$$\begin{aligned} L(w) &= (-1)^2 + 2(1)^2 + w_3 - 2w_3^2 + w_3^2 + 2(-1) - 4(1) + 4 \\ &= 1 + 2 - 2 - 4 + 4 \\ &= 1 \end{aligned}$$

\textcircled{d} No, there is no unique solution w at which this minimum is realized as all points with $w_1 = -1$, $w_2 = 1$, and $w_3 = w_4$ all attain this minimum.



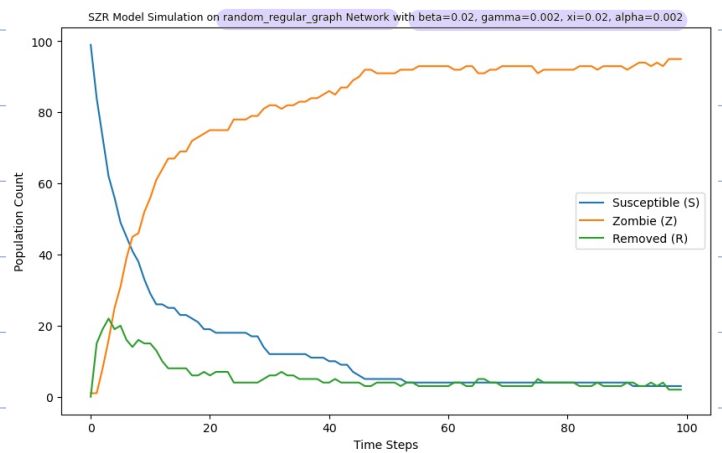
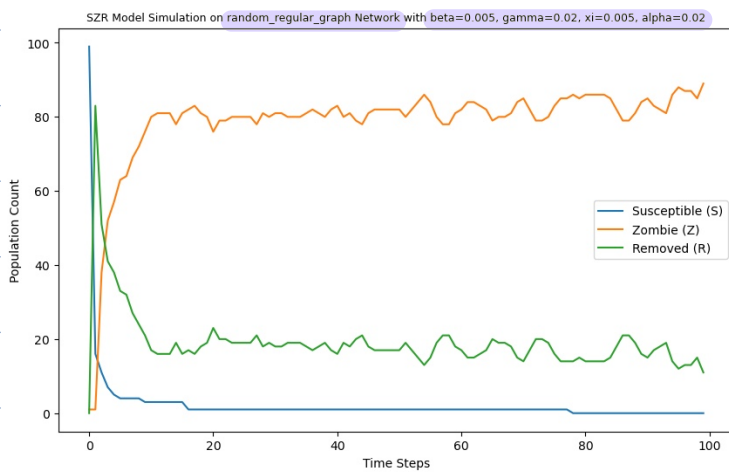
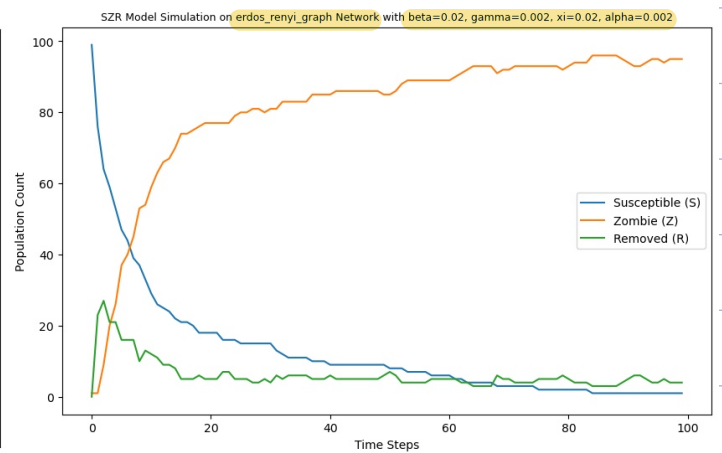
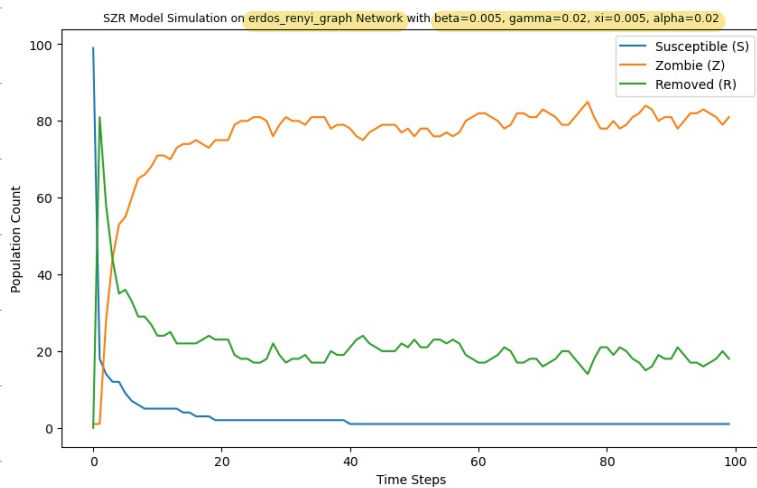
Based on the graph obtained, we can observe that there is a steady increase in the fixation probability as the value of s increases in both "death-birth" and "birth-death" processes.

The fixation probability in the "death-birth" remains slightly lower compared to the fixation probability in the "birth-death" process for all values of s . One reason as to why we observe such similar fixation probabilities in the simulations for "death-birth" and "birth-death" processes is because I used a k -regular graph to create my network.

A k -regular graph has a regular graph structure where each node has exactly k neighbors which creates a well-mixed (more homogeneous) population and therefore minimizes the impact of the order of events. One reason why

we observe a slightly lower fixation probability in the case of "death-birth" is since the death event occurs first, it can potentially remove a "more" fit allele from the population before it has reproduced. In this case, it would take slightly longer for the fitness advantage to demonstrate a spread of "A" alleles in the population.

4.



The S2R model was simulated for the above mentioned parameter values for Erdős-Rényi graph and Random Regular graph. Although I expected to observe a steeper increase in the number of zombies over time in the Erdős-Rényi graph because it is a more densely connected graph compared to the Random Regular graph as a more densely connected graph will display a faster spread of infections. My simulation failed to show such a distinction between the Erdős-Rényi graph and Random Regular graph. Nonetheless, we can see that when ξ and β are given higher values than α , there is a sharper rise in the number of zombies in the population. When the value of α is higher than ξ and β , the number of zombies in the population does not rise up as rapidly. The number of zombies almost reaches the total population, indicating zombification, in the Random-regular graph whereas the number of zombies kind of plateaus at a value slightly lower than total population in the Erdős-Rényi graph.