ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

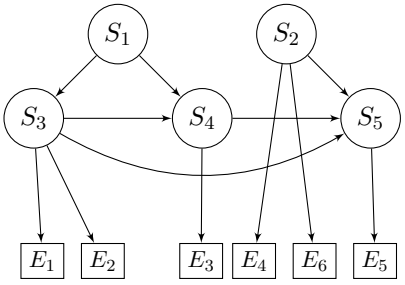
D-BSSE
Department of Biosystems
Science and Engineering

Statistical Models in Computational Biology

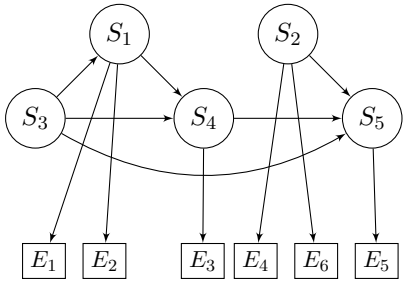
Jack Kuipers
David Dreifuss
Xiang Ge Luo
Rudolf Schill

Due 27th of April 2023
Please submit your project with the filename `Lastname(s)_Project8.pdf`.

Problem 20: Classical NEMs (3 points)



(a)

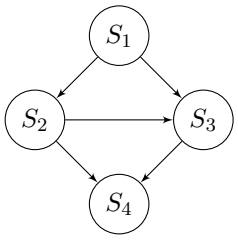


(b)

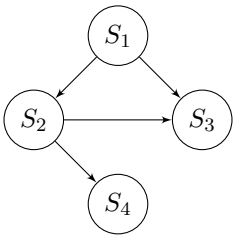
Solve this problem in R.

1. For each model, construct the transitive closure (by adding edges) and define the corresponding adjacency matrices Φ and Θ, which represent the signalling pathways and the E-gene attachments. Determine the corresponding expected effect patterns (F). (1.5 point)
2. Assuming **no noise**, determine the discrete data D_1 and D_2 from both models. Given only the data, can you tell apart the two models? (0.5 point)
3. Use the `mnem`[1] package for this question: Take D_1 and D_2 from the previous question. For each model, calculate the marginal log-likelihood ratio (network score) given the data by setting the false positive rate to be 5% and the false negative rate to be 1%. (1 point)

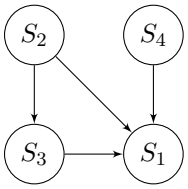
Problem 21: Hidden Markov NEMs (3 points)



u



v_1



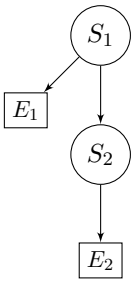
v_2

[1]Please install this version: `https://www.bioconductor.org/packages/devel/bioc/html/mnem.html`

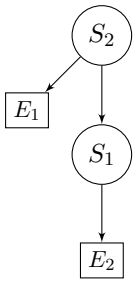
Solve this problem in R.
(*Hint: The `mnem` package does not have an inbuilt function to compute the transition probabilities directly. From the lecture slides implement the different steps and use the functions in the `mnem` package wherever necessary. Please use `mnem::enumerate.models` to enumerate networks.*)

1. Using the definitions for HM-NEMs from the lecture, compute the transition probabilities from $G_t = u$ to $G_{t+1} \in \{v_1, v_2\}$ for different smoothness parameter $\lambda \in \{0.1, \ldots, 0.9\}$. (2 points)
2. Plot the transition probabilities for v_1 and v_2 as a function of λ. Describe the transition probabilities as a function of λ. (1 point)

Problem 22: Mixture NEMs **(4 points)**



F_1



F_2

Given are two NEMs F_1 and F_2 with two S-genes $\{S_1, S_2\}$ and two E-genes $\{E_1, E_2\}$. The data contains four cells $\{C_1, C_2, C_3, C_4\}$. $\{C_1, C_3\}$ are perturbed by a knock-down of S_1, and $\{C_2, C_3, C_4\}$ are perturbed by a knock-down of S_2. (*Hint: You can choose to solve in R or by hand.*)

1. Determine the the cellular perturbation map ρ, where $\rho_{ic} = 1$ if cell c is perturbed by a knock-down of S-gene i. (0.5 points)
2. Assume that $\{C_1, C_2\}$ are generated from F_1 and $\{C_3, C_4\}$ are generated from F_2, compute the **noiseless** log odds matrix R, where $R_{jc} > 0$ means that the perturbation on cell c has an effect on E-gene j:
   (a) For each component k, compute the expected effect pattern $(\rho^T \phi_k \theta_k)^T$. Replace all non-zeros by 1. (1 points)
   (b) Based on the component assignment for each cell, extract the corresponding column from the expected effect patterns computed above and put it into R. Replace all zeros by -1. (0.5 points)
3. Take R from the previous question. Given the vector of mixture weights $\pi = (0.44, 0.56)$, calculate the responsibilities Γ. Then, update the mixture weights. (2 points)