

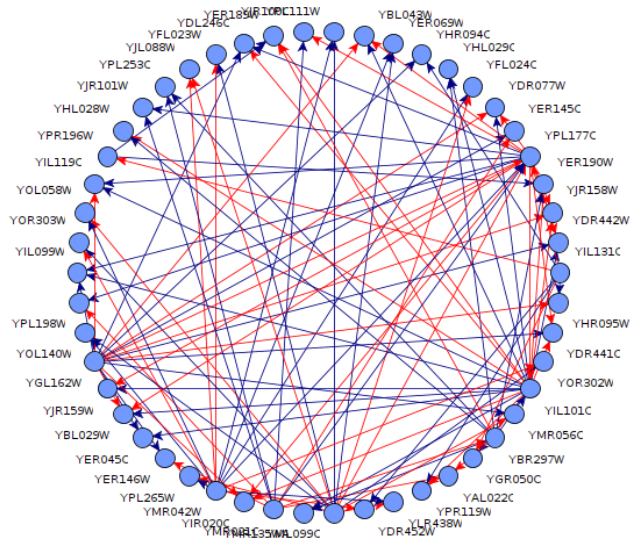
# Linear Gene Networks

Clayton W. Seitz

March 21, 2022

# Outline

## Example gene regulatory network in yeast

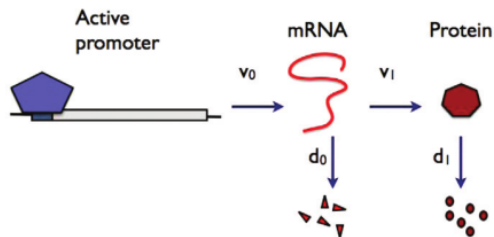


# Linear dynamics of transcription and translation

Assumptions: gene-gene interactions are linear, noise is Gaussian, long protein lifetimes

$$\dot{x}_i = \sum_j m_{ij} y_j - \alpha_i x_i + \eta_i$$

$$\dot{y}_i = r_i x_i - \beta_i y_i$$



If we assume that  $\dot{y}_i \approx 0$  we have a Langevin equation for  $x(t)$  and

$$y/x = \beta/r$$

Let  $\gamma_{ij} = m_{ij}\beta/r$ . An example of a 3-gene system:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -\alpha_1 & \gamma_{21} & \gamma_{31} \\ \gamma_{12} & -\alpha_2 & \gamma_{32} \\ \gamma_{13} & \gamma_{23} & -\alpha_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}$$

# Ornstein-Uhlenbeck process

We have a linear SDE,

$$dx_i = \gamma_{ij}x_j dt + \sigma_{ij}dW$$

which has a corresponding Fokker-Planck equation:

$$\frac{\partial \tilde{P}(\vec{x}, t)}{\partial t} = -\gamma_{ij} \frac{\partial}{\partial x_j} x_i \tilde{P}(\vec{x}, t) + D_{ij} \frac{\partial^2 \tilde{P}(\vec{x}, t)}{\partial x_i \partial x_j} \quad (1)$$

If the real part of the eigenvalues of  $\gamma_{ij}$  are greater than zero, a stationary distribution exists

# Conditional distributions of a Gaussian

Partition variables  $\{x_n\}_{n=1}^N$  into sets  $\mathbf{x}_a$  and  $\mathbf{x}_b$ .

$$\mu = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$$

The conditional distribution  $p(\mathbf{x}_a|\mathbf{x}_b)$  must also be normal with parameters

$$\begin{aligned} \mu_{a|b} &= \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(\mathbf{x}_b - \mu_b) \\ \Sigma_{a|b} &= \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \end{aligned}$$

Bayesian networks loosely express causal relationships. We can compare  $p(\mathbf{x}_a|\mathbf{x}_b)$  and  $p(\mathbf{x}_a)$ . We can use this to assess quality of inference algorithms estimating the underlying network structure parameterized by the damping matrix  $\Gamma_{ij}$

## Marginal distributions of a Gaussian

The conditional distribution  $p(x_1|x_2)$  between two variables  $\mathbf{a} = x_1$ ,  $\mathbf{b} = x_2$  has parameters

$$\begin{aligned}\mu_1 &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \sigma_{1|2}^2 &= \sigma_1^2 - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\end{aligned}$$

The multivariate normal has the nice property that marginal distributions are

$$p(x_1) = \mathcal{N}(\mu_1, \sigma_1^2)$$

Conditional independence implies that  $\mathcal{N}(\mu_1, \sigma_1^2) = \mathcal{N}(\mu_{1|2}, \sigma_{1|2}^2)$ . We can then factor  $p(\mathbf{x})$  into a Bayesian network.

# Handling systems out of equilibrium

In practice we cannot necessarily assume that the data is Gaussian or that the distribution is stationary. It is very difficult to guarantee equilibrium as we can for simulations

We also may need to perform preprocessing on the data to determine if there is population heterogeneity (VAE?)



# Model inference of linear dynamical systems