Goal

(italicized words identify optimization goals)

Wish to find W which has the *most* regular behavior over the *largest* contiguous subset of initial concentrations C_0, given gaussian disturbances of W. The contiguity may not be required, but seems to make sense (modelling a region which the cell can reliably keep the ambient concentrations within). We could take into account non-contiguous subsets, but that is not likely to be useful (it would be weird to have a motif which has similar behavior over different non-contiguous regions of ambient concentration).

Formulation

We define the norm over a matrix.

$$W - W' = W'', \text{ where } W''_{ij} = W_{ij} - W'_{ij} \ \forall \ i,j \tag{1}$$

$$|W| : \mathbb{R}_{n \times n} \to \mathbb{R} = \sqrt{\sum_{i,j \in [1,n]} W_{ij}} \tag{2}$$

Given fixed parameters $W \in \mathbb{R}_{n \times n}, \sigma_W \in \mathbb{R}$, define a Gaussian distribution of weights.

$$D(W'|W, \sigma_W) = (2\sigma_W^2 \pi)^{-2} e^{\frac{-(|W'-W|)^2}{2\sigma_W^2}} \tag{3}$$

This says "mutations modify the weights of the matrix W' a Euclidean distance from the original weight matrix W with probability according to the gaussian distribution with parameter σ_W." This may not be realistic, but it captures the idea that if a mutation occurs, it is likely to modify multiple weights at once if it modifies any at all.

We define the "behavior" of a network (C_0, W) as the concentration vector $C(T)$ at some given time T. This assumes linear differential equations modelling activation and inhibition rates with no external input or output.

$$C_{C_0,W}(t) = C_0 + \int_0^t W \cdot C(t) dt \tag{4}$$

This may not be appropriate; some networks may have *very* reliable periodic behavior, where the period simply changes slightly in response to disturbance, which would lead to vastly different $C(T)$. However, if we restrict ourselves to networks which reach some sort of equilibrium (which need not be defined or used at this time), we can use just $C(T)$. We define networks that reach equilibrium over some region "centered" at C_0 as networks which have similar $C(T)$ for some large T over C'_0 normally distributed with mean C_0. $T > 0$

and must be "large enough to allow the system to get to equilibrium." We can accomplish this by simply choosing a timescale and picking networks which achieve equilibrium.

Variability

Variability V of a network is defined using variation v with an initial concentration (C_0, W) over some gaussian distribution of C_0' is defined.

$$v(C_0, W', W) = \int_{C_0' \in \mathbb{R}^n} D(C_0'|C_0, \sigma_C) \cdot |C_{C_0', W'}(T) - C_{C_0, W}(T)| \tag{5}$$

$$V(C_0, W) = \int_{W' \in \mathbb{R}_{n \times n}} D(W'|W, \sigma_W) \cdot v(C_0, W', W)$$

$$= \int_{W', C_0'} D(W'|W, \sigma_W) D(C_0'|C_0, \sigma_C) \cdot |C_{C_0', W'}(T) - C_{C_0, W}(T)| \tag{6}$$

Finding stable networks X, then, is defined as an optimization problem. The most robust networks (the ones with the minimum variability) are (hopefully) likely to produce interesting network motifs.

$$X^* = \operatorname*{argmin}_{C_0, W} V(C_0, W) \tag{7}$$

Simulation of (C_0, W) seems possible by discretizing choices of C_0' and W' normally distributed about C_0 and W.

Simulation

Use some optimization algorithm to find C_0, W with variance defined as optimal above. Instead of "weighting" the distance $|W' - W|$, $|C_0' - C_0|$, just generate new samples C_0', W' according to normal distribution centered at C_0, W (don't weight these!). These neighboring samples can then be used as neighboring solutions to advance to the next iteration step of an iterative optimization algorithm such as simulated annealing or genetic programming.

Can use iterative refinement for regions which display stability! Also, look for analytical solutions to the PDEs? That would make simulation a lot easier. Can use optimization algorithm like simulated annealing which requires neighboring solutions – since we need to look at neighboring solutions anyway to gauge "variance", we can do this cheaply.