

- Do gene networks drift into unnecessarily complex configurations? If the optimal dynamics of a system can be realized by g genes and/or started in a state with only g genes, during evolution, will the system be composed of $> g$ genes. How many more genes than necessary?
- Are there forces constraining gene network size other than a fitness cost associated with the expense of unnecessary energy? For instance, will genes in a non-minimal gene network contribute less overall to the system's output, diminishing the cost of their removal (on average)?
- as the population explores neutral network space, how often will the population be on or near a point on the manifold where a gene becomes unnecessary? Does this probability go up significantly as unnecessary dimensionality goes up?
- Does the unnecessary complexity of some gene networks confer evolvability advantages?

Let Ω_n be the set of all allowable networks: this is some bounded subset of $\mathbb{R}^{n \times n}$, and let

$$\mathcal{F}_n = \{A : C_n(zI - A)^{-1}B_n = H(z)\} \subseteq \Omega_n$$

be the d_n -dimensional manifold of equivalent systems, where A is gene network (not necessarily minimal), and $H(z)$ is a description of the phenotype in the Laplace domain. In the non-minimal case, C_k and B_k are augmented with zeroes.

$$C_n = \begin{bmatrix} C_k \\ 0 \end{bmatrix}, C_i \neq 0$$

$$B_n = \begin{bmatrix} B_k & | & 0 \end{bmatrix}, B_i \neq 0$$

Let,

$$P_i : \mathcal{F}_n \rightarrow \Omega_{n-1}$$

be the projection from an $n \times n$ network to an $(n-1) \times (n-1)$ network obtained by removing the i^{th} row and column. Biologically, this is a gene deletion or removal from the system.

Let $G_{n,i}$ be the set of networks with identical phenotypes following a deletion.

$$G_{n,i} = \{f \in \mathcal{F}_n : P_i f \in \mathcal{F}_{n-1}\}$$

$$G_n = \cup_i G_{n,i}$$

$$d(f) = \# \{i : P_i f \in \mathcal{F}_{n-1}\}$$

$d(f)$ is the number of different genes that can potentially be deleted to end up in a phenotypically identical space.

Our basic question is: for a given ϵ , how much of \mathcal{F}_n is within ϵ of G_n ? The exact answer will depend on the details, but maybe we see how this changes with n .

One way of getting at this is to look in the other direction. For instance, we know immediately a set of networks in which a gene can be deleted with no effect: it is those networks that just had a gene added. Given a network A we can add a new gene without changing its phenotype by inserting an arbitrary row x into the i^{th} slot and a corresponding column of zeros. We denote this operation $U_i(x)$, and:

$$U_i(x)\mathcal{F}_{n-1} \rightarrow \mathcal{F}_n$$

$$\{U_i(x)f : x \in \mathbb{R}^n, f \in \mathcal{F}_{n-1}\} \subseteq G_{n,i}$$

(Note that gene duplication would do something more specific if we assume the new gene copy does the same thing as the old one.) This implies that $d_n \geq d_{n-1} + n$.

Now we want to answer one of two questions:

How much of a d_n -dimensional manifold is within ε of one of $(n - k)$ sub-manifolds, each of dimension $(d_{n-1} + n)$? This scales like

$$(n - k)\varepsilon^{d_n - d_{n-1} - n}$$

Alternatively, what is the hitting time of $U(x)$?