Recursively constructing analytic expressions for equilibrium distributions of stochastic biochemical reaction networks

Supplementary material

The Journal of the Royal Society Interface

X. Flora Meng*[1,2], Ania-Ariadna Baetica†[3], Vipul Singhal†[4] and Richard M. Murray[3,5]

[1]Mathematical Institute, University of Oxford, Oxford, Oxfordshire, UK
[2]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[3]Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA, USA
[4]Computation and Neural Systems, California Institute of Technology, Pasadena, CA, USA
[5]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA

30 April 2017

Proof of Proposition 1

Proposition 1. *A graph can be obtained by gluing paths together at one vertex sequentially if and only if the graph is a tree.*

Proof. We first show that a graph that can be obtained by gluing paths at one vertex sequentially is a tree. It is apparent that any graph that is obtained by gluing paths at one vertex sequentially is connected. This statement follows from a simple inductive proof on the number of paths that we glue together. Hence, we only need to prove that any graph that is obtained by gluing paths at one vertex sequentially is acyclic. We assume by contradiction that gluing paths $P_1, P_2, \ldots, P_k$ at one vertex sequentially gives a graph G that contains a cycle denoted by C. Then $\{P_1, P_2, \ldots, P_k\} \cap C$ is a set of subpaths and vertices, and cycle C can be obtained by gluing these subpaths at one pair of endvertices (of different subpaths) sequentially. However, this contradicts the fact that gluing two paths at their common endvertex always gives a path. Figure 1 is an example that illustrates the argument.

Conversely, we now show that any tree can be obtained by gluing paths at one vertex sequentially. A tree with one vertex is simply a vertex, and it can be obtained trivially by gluing two 0-paths together. For the inductive step, let n be any positive integer, and suppose that the claim holds for all trees on n or fewer vertices. Consider any tree T with $n+1 \geq 2$ vertices. Recall that any tree with at least two vertices has at least two leaves [1]. Let vertex v be a *leaf* of T (i.e. $d_T(v) = 1$), and let e be the only edge for which v is an endvertex. By hypothesis, T is connected, so any two distinct vertices $x, y \in V(T-e) \subset V(T)$ are connected by a path in T, say P. The vertex v is not an endvertex of P and $d_T(v) = 1$, so vertex v and edge e cannot be on P. Hence, P is a path in $T-e$. Since $T-e \subset T$ and T is acyclic, it follows that $T-e$ is also acyclic. The graph $T-e$ is also connected, so $T-e$ is a tree. But $|T-e| = n$, so by the induction hypothesis $T-e$ can be obtained by gluing paths at one vertex sequentially. But e is a 1-path, and T can be obtained by gluing e and $T-e$ at the only vertex that is adjacent to v. Hence, T can be obtained by gluing paths at one vertex sequentially. By induction, any tree can be obtained by gluing paths at one vertex sequentially. $\square$

*e-mail: xmeng@mit.edu
†Equal contributions.

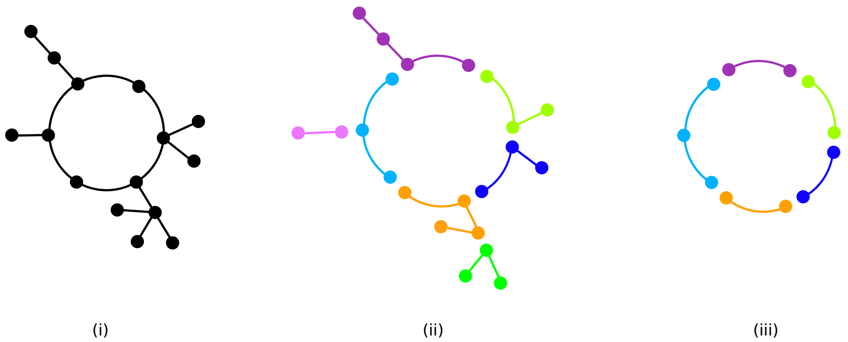

Figure 1: Subfigure (i) is an example of a graph that contains a cycle. If Subfigure (i) could be obtained by gluing paths at one vertex sequentially, as shown in Subfigure (ii), then the cycle could be obtained by gluing subpaths at one vertex sequentially. However, this contradicts the fact that gluing two paths at their common endvertex always gives a path.

Proof of Proposition 2

Proposition 2. *A graph can be obtained by gluing cycles together at one vertex sequentially if and only if the graph satisfies all of the following conditions:*

(i) the graph is connected,

(ii) every vertex has an even degree, and

(iii) any two distinct cycles have at most one common vertex.

Proof. We first show that a graph that can be obtained by gluing cycles at one vertex sequentially satisfies conditions (i)–(iii) by induction on the number of cycles that we use to construct such a graph.

For the base case, consider any cycle. A cycle is connected and every vertex has degree 2. Condition (iii) is trivially true.

For the inductive step, let k be any positive integer and we assume that gluing k or fewer cycles at one vertex sequentially gives a graph that satisfies conditions (i)–(iii). Let G be a graph that is obtained by gluing k cycles at one vertex sequentially. Let C be an arbitrary cycle. Pick any vertex of G and C, say u, and glue G and C at vertex u. We name the new graph $\tilde{G}$. We check that conditions (i)–(iii) hold for $\tilde{G}$.

(i) Since C and G are connected graphs, it follows that $\tilde{G}$ is connected.

(ii) We have $d_{\tilde{G}}(u) = d_G(u) + d_C(u) = d_G(u) + 2$. For all $v \in V(G) \setminus \{u\}$, $d_{\tilde{G}}(u) = d_G(u)$. But every vertex in G has an even degree (in G) and, for all $w \in V(C) \setminus \{u\}$, we have $d_{\tilde{G}}(w) = d_C(w) = 2$. Thus every vertex in $\tilde{G}$ has an even degree (in $\tilde{G}$).

(iii) Consider any two distinct cycles C_1 and C_2 in $\tilde{G}$. If $C_1 = C$ or $C_2 = C$, then C_1 and C_2 have at most one common vertex by the construction of $\tilde{G}$. Otherwise, we have $C_1, C_2 \subset G$, and $|V(C_1) \cap V(C_2)| \leq 1$ by hypothesis. Hence, any pair of distinct cycles in $\tilde{G}$ has at most one common vertex.

By induction, any graph that can be obtained by gluing cycles at one vertex sequentially satisfies conditions (i)–(iii).

We now prove the converse by induction on the number of cycles in a graph that satisfies conditions (i)–(iii).

For the base case, we consider a graph H that contains exactly one cycle $\tilde{C}$ and satisfies conditions (i)–(iii). We assume for a contradiction that H is not a cycle. Then $H - \tilde{C}$ is a *forest* (i.e. an acyclic graph). Every vertex of $H - \tilde{C}$ has degree of the same parity as it does in H. Every *component* (i.e. maximal connected subgraph) of $H - \tilde{C}$ is a tree on at least two vertices, so every component has two leaves. But every leaf has degree 1 by definition. This contradicts our assumption that every vertex in H has an even degree.

For the inductive step, let k be any positive integer, and suppose that any graph that has exactly k or fewer cycles and satisfies conditions (i)–(iii) can be obtained by gluing cycles at one vertex sequentially. Let $\hat{H}$ be a graph that satisfies conditions (i)–(iii) and contains exactly $k+1$ cycles. Pick any cycle $\hat{C} \subseteq \hat{H}$. Without loss of generality, suppose that $\hat{H} - \hat{C}$ consists of components $O_1, O_2, \ldots, O_r$ for some positive integer r. We check that each component O_i $(1 \leq i \leq r)$ satisfies conditions (i)–(iii).

(i) Every component O_i is connected by definition.

(ii) For every $x \in V(O_i) \setminus V(\hat{C})$, $d_{O_i}(x) = d_{\hat{H}}(x)$. For every $y \in V(O_i) \cap V(\hat{C})$, $d_{O_i}(y) = d_{\hat{H}}(y) - 2$. But every vertex in $\hat{H}$ has an even degree. Thus every vertex in O_i has an even degree (in O_i).

(iii) Since $\hat{H}$ satisfies condition (iii) and $O_i \subseteq \hat{H}$, condition (iii) also holds for O_i.

But $O_i \subseteq \hat{H}$ and it has fewer cycles than $\hat{H}$. By hypothesis, O_i can be obtained by gluing cycles at one vertex sequentially.

Since $\hat{H}$ is connected, every component of $\hat{H} - \hat{C}$ shares at least one common vertex with cycle $\hat{C}$. We assume for a contradiction that a component O_j $(1 \leq j \leq r)$ has at least two vertices that are also on $\hat{C}$. Suppose that $\hat{C} = z_1 z_2 \ldots z_s$ for some integer $s \geq 3$. Pick $1 \leq p < q \leq s$ such that z_p, $z_q \in O_j$ and $q - p \in \mathbb{Z}_{>0}$ is a minimal number. Since O_j is connected, there exists a z_p–z_q path (i.e. a path with endvertices z_p and z_q) in O_j, say P. Cycle $\hat{C}$ contains two z_p–z_q paths. If the two paths are of different lengths, then let $\hat{P}$ be the shorter one and we have $V(\hat{P}) \cap V(O_j) = \{z_p, z_q\}$. If the two z_p–z_q paths are of the same length, then $V(\hat{C}) \cap V(O_j) = \{z_p, z_q\}$ and pick either of the z_p–z_q paths as $\hat{P}$. In either case, $V(P) \cap V(\hat{P}) = \{z_p, z_q\}$. By our definition of removing subgraphs from a graph, component O_j has at least 3 vertices. Since edge $z_p z_q \notin O_j$, there must exist some $z \in V(O_j) \setminus \{z_p, z_q\}$ such that $z \in V(P)$. Thus P has length at least 2. But $\hat{P}$ has length at least 1, and $E(P) \cap E(\hat{P}) = \emptyset$. Therefore, $P \cup \hat{P}$ has length at least 3. Hence, $P \cup \hat{P}$ is a cycle in $\hat{H}$ that is different from $\hat{C}$ and $\{z_p, z_q\} \subseteq V(P \cup \hat{P}) \cap V(\hat{C})$. This constradicts our assumption that any pair of distinct cycles in $\hat{H}$ has at most one common vertex.

Hence, every component O_i $(1 \leq i \leq r)$ has exactly one common vertex with cycle $\hat{C}$. But every component O_i can be obtained by gluing cycles at one vertex sequentially. Hence, $\hat{H}$ can be obtained by gluing cycles at one vertex sequentially. By induction, any graph that satisfies conditions (i)–(iii) can be obtained by gluing cycles at one vertex sequentially. □

Proof of Proposition 3

Proposition 3. *A graph can be obtained by gluing paths and cycles together at one vertex sequentially if and only if the graph satisfies both of the following conditions:*

(i) the graph is connected, and

(ii) any two distinct cycles share at most one common vertex.

Proof. We first show that a graph that can be obtained by gluing paths and cycles at one vertex sequentially satisfies conditions (i) and (ii) by induction on the number of paths and cycles that we use to construct such a graph.

We have two base cases: a path and a cycle. In both cases, the graph is connected and condition (ii) holds trivially.

For the inductive step, let m and n be any nonnegative integers such that $m + n \geq 1$, and suppose that any graph that is obtained by gluing m or fewer paths and n or fewer cycles satisfies conditions (i) and (ii). Let G be any graph that is constructed by gluing m paths and n cycles at one vertex sequentially. Let P be any path. We glue P and G at any vertex, say

u, and name the new graph $\tilde{G}$. Since P and G are connected graphs, it follows that $\tilde{G}$ is also connected. Since $|V(P) \cap V(G)| = 1$ and P is a path, no cycle in $\tilde{G}$ contains any edge on P. Thus every cycle in $\tilde{G}$ is a subgraph of G. But any two distinct cycles in G have at most one common vertex (in G). Hence, condition (ii) holds in $\tilde{G}$. Now let C be any cycle. Glue C and G at any vertex, say v. Let $\hat{G}$ be the new graph. Since C and G are connected graphs, it follows that $\hat{G}$ is also connected. Let C_1 and C_2 be any two distinct cycles in $\hat{G}$. If $C_1 = C$ or $C_2 = C$, then $V(C_1) \cap V(C_2) \leq 1$ by the construction of $\hat{G}$. Otherwise, since $V(C) \cap V(G) = 1$, we have $E(C) \cap E(G) = \emptyset$, so $\{C_1, C_2\} \subseteq G$. Therefore, $V(C_1) \cap V(C_2) \leq 1$ by the induction hypothesis.

By induction, any graph that is obtained by gluing paths and cycles at one vertex sequentially satisfies conditions (i) and (ii).

We now prove the converse by induction on the number of cycles in a graph.

A connected graph that contains no cycles is a tree and satisfies condition (ii) trivially. We have proved in Proposition 1 that any tree can be obtained by gluing paths at one vertex sequentially.

For the inductive step, let k be any nonnegative integer, and suppose that any graph that contains exactly k or fewer cycles and satisfies conditions (i) and (ii) can be obtained by gluing paths and cycles at one vertex sequentially. Let H be a graph that satisfies conditions (i) and (ii) and contains exactly $k+1$ cycles. Let C be any cycle in H. Without loss of generality, suppose that $H - C$ consists of components $O_1, O_2, \ldots, O_r$ for some positive integer r. Components are connected by definition. Every component O_i $(1 \leq i \leq r)$ is a subgraph of $H - C$, so O_i has fewer cycles than H and condition (ii) holds for O_i $(1 \leq i \leq r)$. By hypothesis, O_i can be obtained by gluing paths and cycles at one vertex sequentially. Following a similar argument as in the proof for Proposition 2, we can prove that every component O_i has exactly one common vertex with cycle C. Hence, H can be obtained by gluing paths and cycles at one vertex sequentially. By induction, any graph that satisfies conditions (i) and (ii) can be obtained by gluing paths and cycles at one vertex sequentially. $\square$

A graphic illustration of truncating infinite state spaces

Figure 2 illustrates how to truncate the infinite state space of two interconnected transcriptional components to a finite subset when $N = 3$ and $M = 2$.

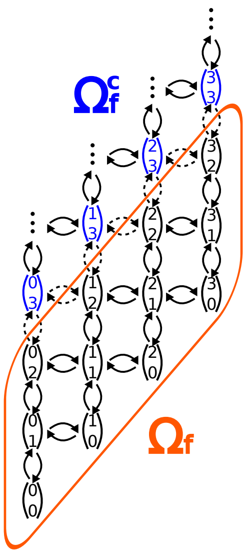

Figure 2: A graphic illustration of truncating the infinite state space of two interconnected transcriptional components to a finite subset when $N = 3$ and $M = 2$. The truncated finite state space, Ω_f, lies within the orange curve. We highlight in blue the first layer of the complement infinite state space, $\Omega_\mathrm{f}^\mathrm{c}$. The transition edges between Ω_f and $\Omega_\mathrm{f}^\mathrm{c}$ are represented by dashed arrows, the transition rates of which determine the probabilities of Ω_f and $\Omega_\mathrm{f}^\mathrm{c}$ at equilibrium.

Proof of Proposition 4

Proposition 4. *Consider the system of two interconnected transcriptional components that are modelled by reactions as given in Equation* (12)*, where* $\kappa > 0$*,* $\delta > 0$*,* $\kappa_{on} > 0$*, and* $\kappa_{off} > 0$ *are the corresponding reaction rate constants. Let* P*,* Z*, and* C *be the numbers of promoters* $\mathcal{P}$*, transcription factors* $\mathcal{Z}$*, and* $\mathcal{P}$*–*$\mathcal{Z}$ *complexes* $\mathcal{C}$*, respectively. Let* $\alpha = \frac{\kappa\kappa_{on}}{\delta\kappa_{off}}$*,* $\beta = \frac{\kappa}{\delta}$*, and* $\gamma = \frac{N\alpha-1}{\alpha+1}$*, where* N *is a constant given by* $N = P + C$ *due to the conservation of DNA. In (i)–(iii), we set up and solve three design problems using the marginal stationary distributions of* Z *and* C*.*

(i) Since the marginal stationary distribution of Z *is Poisson distributed, its mean and variance are equal. The design problem of fixing the mean of* Z *at an objective value* $\mu_z > 0$ *is feasible, and the solution is* $\beta = \mu_z$*.*

(ii) The design problem of setting the mean of C *at an objective value* $\mu_c \in (0, N)$ *is feasible, and the solution is* $\alpha = \frac{\mu_c}{N-\mu_c}$*.*

(iii) The design problem of choosing the variance of C *to be an objective value* $\sigma_c^2 > 0$ *is feasible if and only if* $\sigma_c^2 \le \frac{N}{4}$*, and the solutions are* $\alpha = \frac{N-2\sigma_c^2 \pm \sqrt{N^2-4N\sigma_c^2}}{2\sigma_c^2}$*.*

Proof. The marginal stationary distribution of Z is Poisson distributed with mean and variance equal to β. Setting the mean and variance of Z at $\mu_z > 0$ is equivalent to specifying $\beta = \mu_z$, which is always feasible. The marginal stationary distribution of C is binomially distributed with the number of trials and success probability in each trial being N and $\frac{\alpha}{1+\alpha}$, respectively. Design problem (ii) corresponds to setting $\frac{N\alpha}{1+\alpha} = \mu_c$. A solution exists if and only if $0 < \mu_c < N$, in which case the solution is $\alpha = \frac{\mu_c}{N-\mu_c}$. Design problem (iii) is equivalent to solving $\frac{N\alpha}{(1+\alpha)^2} = \sigma_c^2$, leading to $\alpha = \frac{N-2\sigma_c^2 \pm \sqrt{N^2-4N\sigma_c^2}}{2\sigma_c^2}$ for $0 < \sigma_c^2 \le \frac{N}{4}$. □

Proof of Proposition 5

Proposition 5. *Consider the system of two interconnected transcriptional components that are modelled by the reactions in Equation* (12)*. With the same notation as in Proposition 4, the stationary distribution in Equation* (17) *has a unique global maximum if and only if* $N > 1$*,* $\beta > 1$*,* $0 < \gamma < N - 1$*, and* $\beta, \gamma \notin \mathbb{Z}$*. In this case, the maximum is at* $(c^*, z^*) = (\lfloor\gamma\rfloor + 1, \lfloor\beta\rfloor)$*.*

Proof. Let $\|\cdot\|$ denote the l^1-norm on $\mathbb{R}^2$. For $x \ge 0$, let $\lfloor x \rfloor$ denote the integer part of x. Since N is the total number of promoters $\mathcal{P}$ and complex molecules $\mathcal{C}$, it is reasonable to assume that $N > 1$. The sample space of the probability mass function given by Equation (17) is $\Omega = \{(c, z) \mid c = 0, 1, \ldots, N \text{ and } z \in \mathbb{Z}_{\ge 0}\}$. Let Ω_∂ denote the boundary of Ω, namely, $\Omega_\partial = (\{0, N\} \times \mathbb{Z}_{\ge 0}) \cup (\{0, 1, \ldots, N\} \times 0)$. Equation (17) has a strict local maximum at $(c^*, z^*) \in \Omega \setminus \Omega_\partial$ if and only if, for all $(c, z) \in \Omega$ with $\|(c - c^*, z - z^*)\| \le 1$, we have $\Pr(c, z) < \Pr(c^*, z^*)$. Solving these inequalities simultaneously gives $\gamma < c^* < \gamma + 1$ and $\beta - 1 < z^* < \beta$. Since $(c^*, z^*) \in \Omega \setminus \Omega_\partial$, a unique solution exists if and only if $N > 1$, $\beta > 1$, $0 < \gamma < N - 1$, and $\beta, \gamma \notin \mathbb{Z}$. When these conditions hold, the unique strict local maximum on $\Omega \setminus \Omega_\partial$ is at $(c^*, z^*) = (\lfloor\gamma\rfloor + 1, \lfloor\beta\rfloor)$. Moreover, basic algebra shows that, for all $c = 0, 1, \ldots, N$, we have $\Pr(c, z) < \Pr(c, z + 1)$ for all $z = 0, 1, \ldots, z^* - 1$ and $\Pr(c, z) > \Pr(c, z + 1)$ for all integers $z \ge z^*$. It is also straightforward to verify that $\Pr(c, z^*) < \Pr(c+1, z^*)$ for all integers $c = 0, 1, \ldots, c^* - 1$, and $\Pr(c, z^*) > \Pr(c+1, z^*)$ for all integers $c = c^*, c^* + 1, \ldots, N$. Therefore, the stationary distribution of the two-component transcriptional system has a unique global maximum at (c^*, z^*) if and only if $N > 1$, $\beta > 1$, $0 < \gamma < N - 1$, and $\beta, \gamma \notin \mathbb{Z}$. □

References

[1] West DB. 2001 *Introduction to graph theory*, 2nd edn. Upper Saddle River, NJ: Prentice Hall.