Structural determinants of relaxation dynamics in chemical reaction networks

Yusuke Himeoka[a,b], **Julius B. Kirkegaard**[b], **Namiko Mitarai**[b], **and Sandeep Krishna**[c]

[a]Universal Biology Institute, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku , Tokyo, Japan; [b]The Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, Copenhagen, 2100, Denmark; [c]Simons Centre for the Study of Living Machines, National Centre for Biological Sciences TIFR, GKVK Campus, Bellary Road, Bangalore 560065, India

This manuscript was compiled on September 19, 2022

Understanding the relationship between the structure of chemical reaction networks and their reaction dynamics is essential for unveiling the design principles of living organisms. However, while some network-structural features are known to relate to the steady-state characteristics of chemical reaction networks, mathematical frameworks describing the links between out-of-steady-state dynamics and network structure are still underdeveloped. Here, we characterize the out-of-steady-state behavior of a class of artifical chemical reaction networks consisting of the ligation and splitting reactions of polymers and . Within this class we examine minimal networks that can convert a given set of inputs (e.g., nutrients) to a specified set of targets (e.g., biomass precursors). We find three distinct types of relaxation dynamics after perturbation from a steady-state: exponential-, power-law-, and plateau-dominated. We show that we can predict this out-of-steady-state dynamical behavior from just three features computed from the network's stoichiometric matrix, namely, (i) the rank gap, determining the existence of a steady-state; (ii) the left null-space, being related to conserved quantities in the dynamics; and (iii) the stoichiometric cone, dictating the range of achievable chemical concentrations. We further demonstrate that these three quantities also predict the type of relaxation dynamics of combinations of our minimal networks, larger networks with many redundant pathways, and a real example of a metabolic network. The unified method to predict the qualitative features of the relaxation dynamics presented here can provide a basis for understanding the design of metabolic reaction networks as well as industrially useful chemical production pathways.

chemical reaction| network | dynamics | metabolism

Origins of chemical reaction network theory can be found in the discovery of the law of mass action (1). Since then, mathematical laws governing chemical reactions have been uncovered, such as Le Chatelier's principle (2), the Arrhenius equation (3), and the detailed-balance condition for the feasibility of chemical equilibrium (4).

For the development of quantitative theories of living entities, a deep understanding of the out-of-steady-state dynamics of chemical reactions is indispensable. Until recently, most experimental studies of microbial physiology focused on steadily-growing cellular populations. In such situations, the assumption that the concentrations of cellular metabolites are constant (at steady-state) is reasonable, and static metabolic analysis such as Flux Balance Analysis (5–10) and steady-state cell models (11, 12) have provided a useful mathematical description of stable cell growth.

However, recent experiments have been unveiling the dynamic nature of cellular physiology more quantitatively. For instance, in starved *Escherichia coli*, the lag time — the duration that the cell takes to restart growth after the substrate replenishment — was shown to depend on how long the cells are starved (13–17) and the death rates of the starved cells differ depending on the previous culture conditions, even though the starvation condition is identical (18). Such "memory" is a unique feature of out-of-steady-state systems. In addition, the time scales of the memories are much longer than that in steadily-growing cells. It is therefore desirable to develop theoretical frameworks to understand how such slow relaxation dynamics and long-term memory emerge from the structure of chemical networks.

There are two broad types of approaches one might take to explore the link between the network structure and the relaxation characteristics of chemical networks. One is to examine random networks with structural characteristics (such as density of links and degree distribution) identical to natural chemical networks like metabolic networks (19–22). This approach enables us to explore a large ensemble of network structures, but due to the simplification into a random network, it may miss crucial properties of real chemical networks. For example, the central principle of chemical reactions, the law of mass conservation, may be violated. Thus, any structure-dynamics connections uncovered by these approaches may not always apply to real chemical networks.

On the other hand, data-driven approaches such as reaction pathway prediction by using machine learning (23–25) generate realistic chemical reaction networks, which never violate mass conservation and other necessary properties. However, it is too computationally costly to explore a wide variety

Significance Statement

The relationship between network structure and chemical reaction dynamics is of central interest in chemical reaction network theory, as it underlies chemical manufacturing, cellular metabolism, and bioengineering. The links between structure and steady-state properties have been extensively investigated. However, how far the network structure determines the out-of-steady-state, transient dynamics of chemical reactions is unexplored. Here we construct a chemical reaction network model that is simple but generates a wide variety of network instances. By exploring the networks' structural- and dynamical features, we found that three network-structural features are sufficient to predict the qualitative characteristics of the relaxation dynamics after the chemical concentrations are perturbed from their steady-state. Our findings provide a foundation for the structure-based prediction of chemical reaction dynamics.

Please provide details of author contributions here.

Please declare any competing interests here.

[1]A.O.(Author One) contributed equally to this work with A.T. (Author Two) (remove if not applicable).

[2]To whom correspondence should be addressed. E-mail: author.twoemail.com

DRAFT

of possible chemical reaction networks. The generality of structure-dynamics connections, or lack of it, uncovered by these approaches may therefore be unclear.

Here we develop an alternative approach. We construct a class of the chemical reaction network consisting of chemical species that can be represented as monomer sequences (polymers), and chemical reactions that ligate (concatenate) and split (fragment) these polymers. The model can generate a wide variety of chemical reaction network structures that are guaranteed to satisfy the mass conservation principle.

Microbial metabolic networks and chemical networks used in industrial manufacturing processes have the specific function of converting certain input chemicals into target chemicals. For instance, in bacterial metabolism, carbon, nitrogen, phosphorous sources (e.g., glucose, potassium hydrogen phosphate, ammonium chloride) and other miscellaneous molecules are converted into biomass such as proteins, nucleic acids, and lipids. Therefore, we impose a further constraint on the networks we study - the ability to synthesize a given set of targets from a specified set of input chemicals. The resultant chemical reaction networks we generate are, of course, a subset of all possible networks, but they form a well-defined subset so the generality of the structure-dynamics relationships we find is also well-defined. Our framework can also easily be extended to include other kinds of chemical species and reactions as desired.

We find that for these artificial reaction networks we construct, there is a single stable steady-state for the chemical concentrations. Computational simulations of the chemical reaction dynamics upon perturbation of these steady-states revealed four distinct types of relaxation dynamics. Interestingly, we found that the qualitative features of the relaxation dynamics are determined by only three properties of the stoichiometric matrix which encodes the chemical network structure, namely, (i) the rank gap, (ii) the dimension of the left null space, and (iii) the stoichiometric cone. We will define all these properties precisely in subsequent sections, but crudely, a non-zero rank gap indicates that no steady-state exists in the absence of any degradation of chemicals, and the left null space determines the existence of conserved quantities in the dynamics. The stoichiometric cone determines the reachability of the final steady state. Finally, we tested the applicability of these three network-structural features to the prediction of relaxation dynamics of a real biological (metabolic) network. Our results and the general framework we have developed provides valuable tools for exploring the relationship between the network structure and chemical reaction dynamics, which may be applied both to complex industrial manufacturing processes as well as design of synthetic microorganisms designed to perform certain chemical tasks (26, 27).

Model

Constructing a class of artificial reaction networks. For the construction of the chemical reaction networks, we define a set of chemicals $\mathscr{C}$ and the set of the reactions $\mathscr{R}$ they participate in. In our artificial reaction network model, we consider each chemical species to be a one-dimensional chain of the monomers, such as AA or $AABA$ where A and B represent different monomer species. Each monomer idealizes chemical elements such as hydrogen, carbon, nitrogen, and oxygen, or even larger chemical moeities. The chemical reactions possible are the splitting or ligation of the chemicals. For instance, there are three splitting reactions possible with $AABA$ as the substrate: $AABA \to A + ABA$, $AABA \to AA + BA$, and $AABA \to AAB + A$. In the present manuscript, we suppose that all reactions are reversible. That is, if there is a splitting reaction $AABA \to A + ABA$, the corresponding ligation reaction $A + ABA \to AABA$ exists.

With this set-up, the set of all possible chemicals $\mathscr{C}$ and the set of all possible reactions $\mathscr{R}$ are fully determined by setting the number of monomers M and the length of the longest polymer L. For instance, in the case of $M = 2$ and $L = 2$, $\mathscr{C}$ and $\mathscr{R}$ are given by

$$\begin{aligned} \mathscr{C} &= \{A, B, AA, AB, BB\}, \\ \mathscr{R} &= \{AA \rightleftharpoons 2A, AB \rightleftharpoons A + B, BB \rightleftharpoons 2B\}. \end{aligned}$$

To avoid redundancy, we omit the direction of the polymers, e.g., the polymer BA is identified as identical to AB. Given this, the number of chemicals ($|\mathscr{C}|$) and the number of reactions ($|\mathscr{R}|$) scale with the parameters M and L as

$$|\mathscr{C}| \sim \sum_{l=1}^{L} M^l \sim M^L, \ |\mathscr{R}| \sim \sum_{l=1}^{L} (l-1) M^l \sim L M^L.$$

We now construct individual chemical reaction networks by choosing appropriate subsets of $\mathscr{R}$. Importantly, as each reaction in $\mathscr{R}$ satisfies the mass-conservation principle, all the reaction networks made by combining the reactions from $\mathscr{R}$ satisfy this principle. The number of possible reaction networks formed by choosing a subset of $\mathscr{R}$, including disconnected ones, is roughly given by $2^{|\mathscr{R}|} \sim 2^{LM^L}$. This number grows astonishingly fast with M and L (256 for $M = 2$ and $L = 2$, but $\approx 10^{19}$ for $M = 2$ and $L = 4$). Thus, we focus only on the *minimum* networks which can produce a given set of N_{tgt} target chemicals from a given set of N_{in} input chemicals (in metabolic terms, those that can produce biomass from the available nutrients) using the smallest possible number of reactions.

We define a *globally* minimum network to be the set of reactions $\mathscr{R}_0 \subset \mathscr{R}$ with the smallest $|\mathscr{R}_0|$ which can synthesize all the targets from the input chemicals. In contrast a *locally minimum* network is one from which no reaction can be removed while still maintaining production of all targets from the input chemicals. There are many locally minimum networks where the number of reactions $|\mathscr{R}_0|$ is not the global minimum. Note also that we must choose the input chemicals and targets such that the network consisting of all reactions in $\mathscr{R}$ satisfies the condition of being able to synthesize the targets from the input chemicals, otherwise no subset will be able to do so. However, this 'fully connected network' is typically very far from being locally minimum, let alone globally minimum. Finding globally minimum reaction networks, where $|\mathscr{R}_0|$ is the global minimum, is computationally challenging. Here we exploit the fact that this computation can be represented in the form of the 3-satisfiability (3-SAT) problem. Being an NP-complete problem, no polynomial time algorithm exists (28). However, sophisticated 3-SAT solvers are available which work very well in practice — in particular for the present problem for the minimum networks (see Materials and Methods for more details). Henceforth, we will use the term 'minimum network' to mean 'globally minimum network', and explicitly specify it if we mean a 'locally minimum network'.

DRAFT

In the following, we deal with the minimum networks generated from the fully-connected network with $M = 2$ and $L = 6$, with $N_{\rm in} = 2$ and $N_{\rm tgt} = 1$. The fully-connected network and examples of minimum networks are depicted in Fig.1 A and B, respectively.

Defining the chemical reaction dynamics of the constructed networks. There are multiple choices for implementing the kinetics of a given chemical reaction network. In the present manuscript, we utilize the simplest and widely used formulation — mass-action kinetics. Consider the following reaction with n_s and n_p reactants of the forward- and backward reaction, respectively.

$$\sum_{i=1}^{n_s-1} \sigma^+_{s(i)} X_{s(i)} \rightleftharpoons \sum_{i=0}^{n_p-1} \sigma^-_{p(i)} X_{p(i)}.$$

By using mass-action kinetics, the rate of the reaction (also called the 'flux') is given by

$$J = v_+ \prod_{i=0}^{n_s-1} x_{s(i)}^{\sigma^+_{s(i)}} - v_- \prod_{i=0}^{n_p-1} x_{p(i)}^{\sigma^-_{p(i)}}, \qquad [1]$$

where the capital letter X_i indicates the ith chemical, and its concentration is denoted by the corresponding lower case letter. $s(i)$ and $p(i)$ represent the index of the ith reactant of the forward- and backward reactions, respectively. $\sigma_i^\pm$ specify the stoichiometric coefficients of the ith chemical in the forward- (σ_i^+) and backward (σ_i^-) reaction, respectively. $v_\pm$ indicates the rate of the forward (v_+) and the backward (v_-) reaction. When we summarize the information of the whole reaction by using a reaction stoichiometric matrix, the relationship between $\sigma_{ir}^\pm$ (now we added the index of the reaction r) and the stoichiometric matrix S is given by

$$\sigma^+_{ir} = \begin{cases} -S_{ir} & (S_{ir} < 0) \\ 0 & (\text{otherwise}), \end{cases} \qquad [2]$$

$$\sigma^-_{ir} = \begin{cases} S_{ir} & (S_{ir} > 0) \\ 0 & (\text{otherwise}). \end{cases} \qquad [3]$$

Each (minimum) reaction network is represented by a corresponding reaction stoichiometric matrix S_0, which denotes the relationship of the interconversion among the chemicals by the splitting and ligation reactions. The exponents in the mass-action kinetics (Eq. (1)), $\sigma_{ir}^\pm$'s, are set by S_0 according to Eq. (2) and 3. This matrix however does not represent the processes of uptake of input chemicals and harvest of the targets. Therefore, we add the uptake reaction $\tilde{X}_i^{(\rm in)} \rightleftharpoons X_i^{(\rm in)}$ for $N_{\rm in}$ inputs, where the symbol with tilde represent the corresponding chemical in the external environment. Also, we suppose that the target chemicals are harvested together, and once harvested, the target chemicals never returns back to the reaction system, i.e, $\sum_{i=0}^{N_{\rm tgt}-1} X_i^{(\rm prp)} \to \emptyset$. The kinetics of each of these reactions are also implemented by the mass-action kinetics given by

$$J_i^{(\rm in)} = u_+^{(i)} \tilde{x}_i^{(\rm in)} - u_-^{(i)} x_i^{(\rm in)}, \ (0 \le i < N_{\rm in}) \qquad [4]$$

$$J^{(\rm tgt)} = p_+ \prod_{i=0}^{N_{\rm tgt}-1} x_i^{(\rm prp)} - p_- \tilde{x}^{(\rm pr)} \qquad [5]$$

We further suppose that all chemical species are spontaneously degraded or deactivated at a uniform rate ϕ which is sufficiently smaller than the rates of the other chemical reactions. Additionally, the external environment is large enough so the concentrations of chemicals there are set to be constant. Then, by summing up, the dynamics of the concentrations of chemical species follow the equation below

$$\dot{\boldsymbol{x}} = S_0 \boldsymbol{v}_0 + \boldsymbol{u} - \boldsymbol{p} - \phi \boldsymbol{x}, \qquad [6]$$

where $\dot{\boldsymbol{x}}$ represents the time-derivative of $\boldsymbol{x}$, $d\boldsymbol{x}/dt$. S_0 is the stoichiometric matrix of the network. $\boldsymbol{v}_0$ represents the reaction flux vector and each element has the form of Eq. (1). $\boldsymbol{u}$ and $\boldsymbol{p}$ denote the vectors representing the rate of input chemical uptake and target harvesting, respectively. If the ith chemical X_i is one of the inputs (target), the ith element of the vector $\boldsymbol{u}$ ($\boldsymbol{p}$) is given by the nonzero entity $J_i^{(\rm in)}$ ($J^{(\rm tgt)}$).

Once a reaction network and the inputs and targets are specified, the differential equation (Eq. (6)) is fully and uniquely determined. In this sense, we call the Eq. (6) the accompanying ordinary differential equation (ODE) of the reaction network. For simplicity, we set all the rate constants except ϕ to unity.

Relaxation behaviours and network structures

In the following sections, we study the link between the network structure and the relaxation behaviour of the accompanying ODE. First, we generate ca. 2,000 minimum reaction networks (see Methods for the parameters we used) and check if there is a steady-state attractor. As far as we have confirmed, all the networks each have a unique steady-state attractor.

For probing the relaxation characteristics, we applied perturbations to the attractor. Since the concentrations of the chemicals span a wide range at the attractors, we apply perturbations in a multiplicative manner. In addition, to normalize the strength of the perturbations, we generated the initial points on the surface of the $|\mathscr{C}_0|$-dimensional hypersphere, centred at the steady-state attractor, with a given diameter D. In concrete terms, a single initial point $\boldsymbol{x}_{\rm ini}$ is given by $\ln \boldsymbol{x}_{\rm ini} = \ln \boldsymbol{x}_{\rm att} + D \cdot \boldsymbol{r}/\|\boldsymbol{r}\|$ with $\boldsymbol{x}_{\rm att}$ is the steady-state attractor of the model, and $\boldsymbol{r}$ is a vector of the random numbers each lying in the interval $[0, 1]$, respectively (see Materials and Methods section for more details).

For a sufficiently small value of D, such that the dynamics could be described adequately by linearizing the ODE around the steady-state, it is known that the relaxation dynamics would be exponential, with a rate determined by the eignevalues of the Jacobi matrix. Hence, we choose a value of D that is large enough to go beyond this linear regime so we can explore how the non-linearities of the ODEs result in non-exponential relaxation.

We simulated the accompanying ODE (Eq. (6)) of ca. 2,000 minimum networks with $N_{\rm ptb}(= 256)$ initial points for each network. Visually, we found three typical types of relaxation: exponential, plateau, and power-law type. Examples of each are shown in Figs.1C-F. There appeared to be three groups in terms of classification of the networks: (i) the networks exhibiting only exponential relaxation regardless of the initial concentrations, (ii) the networks showing only the plateau type relaxation regardless of the initial concentrations, and (iii) the networks displaying either the plateau or the power-law type of relaxation depending on the initial concentrations. At this

DRAFT

stage, we do not know if the plateau exhibited by the second and the third type of networks are different in any sense, but for later arguments we refer to them as "Metastable Plateau" (independent of initial conditions) and "Confined Plateau" * (dependent on initial conditions) in Figs. 1D and E.

To make the classification more quantitative, we computed two quantities: the relaxation time and the migration length. Here we define the relaxation time as the time point at which the Euclidean distance, in logarithmic scale, between the state and the steady-state attractor first becomes smaller than a certain threshold value ϵ, that is

$$T_{\mathrm{relax}} = \min\{t \mid \| \ln \boldsymbol{x}(t) - \ln \boldsymbol{x}_{att} \| < \epsilon\}. \quad [7]$$

While we adopted $\epsilon = 10^{-6}$, it is checked that our conclusions are unchanged as long as ϵ value is in the range $(10^{-6}, 10^{-2})$.

Next, to distinguish between the plateau and the power-law relaxation types, we quantified the total length of the trajectory in concentration space during a given time period. It is measured by the "migration length" defined as

$$L = \int_a^b \left\| \frac{\mathrm{d} \ln \boldsymbol{x}}{\mathrm{d}t} \right\| dt, \quad [8]$$

where we choose $a = 100$ and $b = 0.01/\phi$ in order to quantify the "motion" of the trajectory in the intermediate time scale (sufficiently larger than the time scale of the reaction rate constants, which are unity, and sufficiently smaller than that of the degradation, $1/\phi$). The migration length is computed in the logarithmic scale to take the effect of low-concentration chemicals into account. Additionally, we compute $L_{\max}$ given by Eq. (8) with $b = 10^{10}$ (the end of the simulation) as the upper bound of the integral.

Our prior expectation, from visual examination of the trajectories (e.g., Figs.1C-F), was that since plateau type trajectories are almost static in the time range $t \in [100, 0.01/\phi]$, the normalized migration length $L/L_{\max}$ would have a relatively lower value, compared to that for power-law type dynamics. We also expected that exponential relaxation dynamics will result in lower T_{relax} values than the other two types.

Examples of the distribution of migration length of two sets of networks are presented in Fig. 1G. The upper panel is the distribution over all initial conditions for one network that shows Metastable-plateau dynamics, and the bottom panel is for one network exhibiting both confined-plateau and power-law relaxation. The bottom distribution shows a clear double-peak due to bimodality of the relaxation dynamics, while the upper distribution is rather unimodal. We suppose that if networks exhibit a bimodal distribution due to the presence of both power-law relaxation in addition to confined-plateau relaxation, the average value of the normalized migration length $L/L_{\max}$ (averaged over all initial conditions) would be larger than the corresponding value for metastable-plateau networks. Thus, we utilize $\langle L/L_{\max} \rangle$ as an indicator of the variety of the relaxation behaviour.

The average relaxation time and the average migration length are computed for each network, and we plotted the distribution over the networks. As shown in Fig.1H, the distribution has four distinct peaks, and we classified the networks from type-I to type-IV according to the nearest peak. While the separation of the type-II and type-IV peaks is less clear, the distribution-based classification of the networks matched very well to the visual grouping of the models: the type-I networks (small T_{relax} and large $L/L_{\max}$) exhibit only exponential relaxation; most of the type-II networks showed metastable-plateau relaxation; and, the type-III networks and most type-IV networks exhibited both the confined-plateau and the power-law type relaxation depending on the initial concentrations.

Note that the final steady-state attractor is reached at the latest by $t \sim 1/\phi$, which is the timescale of the spontaneous degradation, and the qualitative differences between the different types of relaxation only appear in the $t \ll 1/\phi$ region. These observations mean that the difference in the relaxation behaviour during $t \ll 1/\phi$ highlight structural features of the reaction networks and would not be changed by varying the value of $1/\phi$ as long as it is sufficiently large. What structural features of the networks result in the four different relaxation types observed in Fig.1H? We now show that the relaxation behaviour is indeed determined by three features of the stoichiometric matrix (which in turn encodes the structure of the network), namely, (I) the rank gap, (II) dimension of the left null-space, and (III) the stoichiometric cone.

I. The rank gap. The distinct relaxation behaviours are observed over timescales shorter than $1/\phi$, and thus, should stem from the dynamics (Eq. (6)) without the degradation term. The clear difference between the exponential relaxation and the others is the relaxation time. The exponential type trajectories already reach the attractor well before t exceeds $1/\phi$. We now show how measuring the ranks of the stoichiometric matrix and some enlargements can inform us about the reachability to the attractor in the absence of the degradation term.

First, we introduce additional notation. Let S_1 denote the enlarged matrix $(S_0|\boldsymbol{U}^{(0)}|\boldsymbol{U}^{(1)}|\cdots|\boldsymbol{U}^{(N_n-1)})$, where $\boldsymbol{U}^{(j)}$'s are vectors of the stoichiometry of the input uptake reactions, i.e., $U_i^{(j)}$ is unity if the ith chemical is the jth input chemical and zero otherwise. By vertically stacking $\boldsymbol{v}_0$ and $\boldsymbol{u}$ in Eq. (6) and denoting it as $\boldsymbol{v}_1$, we can rewrite Eq. (6) with S_1 instead of S_0. Then, removing the degradation term, we obtain

$$\dot{\boldsymbol{x}} = S_1 \boldsymbol{v}_1 - \boldsymbol{p}. \quad [9]$$

Note that Eq. (9) represents the dynamics of chemical concentration driven by the input chemical uptake, internal reactions, and the target synthesis. This equation approximates the original equation Eq. (6) for $t \ll 1/\phi$.

For there to exist a steady-state of this equation, there must be a pair of vectors, $\boldsymbol{v}_1^*$ and $\boldsymbol{p}^*$, satisfying $S_1 \boldsymbol{v}_1^* = \boldsymbol{p}^*$. Since we are not interested in the trivial solution $\boldsymbol{v}_1^* = \boldsymbol{0}$, the necessary and sufficient condition for the existence of a steady-state is that the rows of S_1 and $\boldsymbol{P}$ (The stoichiometric part of the harvest reaction) are linearly dependent, i.e, $\mathrm{rank} S_1 = \mathrm{rank}(S_1|\boldsymbol{P})$, where $\boldsymbol{P}$ is the stoichiometry of the target synthesis (i.e., the ith element is unity if the ith chemical is one of the target precursors and zero for the rest).

We found that the rank gap $\delta \equiv \mathrm{rank}(S_1|\boldsymbol{P}) - \mathrm{rank} S_1$ for all type-I networks (showing only the exponential relaxation) is zero. Interestingly, all type-II networks also satisfied the

* "Confined" here means two things: the first meaning is that the initial conditions leading to the plateau behaviour are confined in a certain region in the phase space. The second meaning is the confinement of the trajectory in the so-called "stoichiometric cone". See the subsection III for more details.

DRAFT

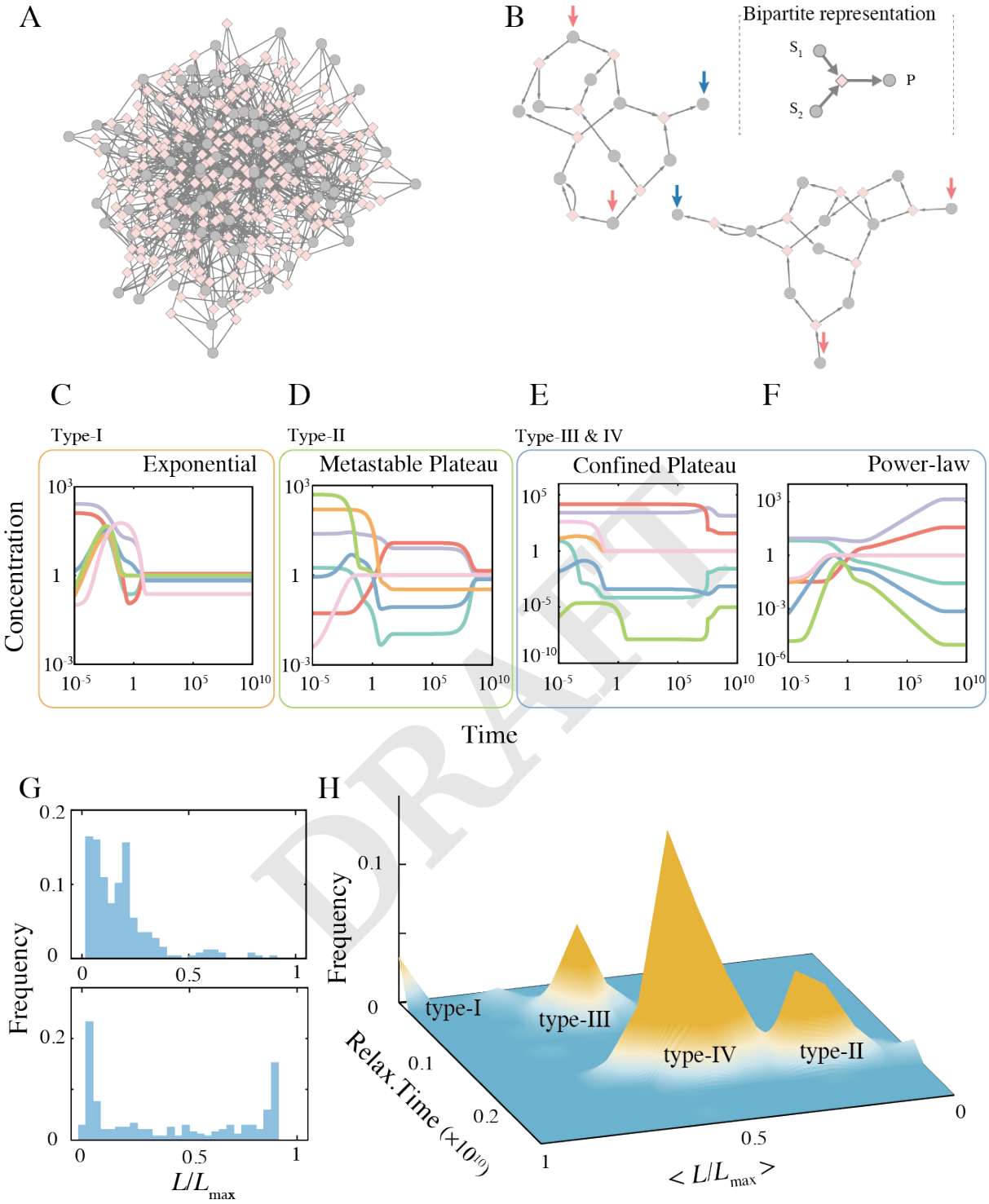

Fig. 1. (A) The fully-connected network for $M = 2$ and $L = 6$. (B) Two examples of the minimum network which synthesizes a single target from two inputs. The inputs and target chemicals are indicated by the red arrows and the blue arrow, respectively. In A and B, the bipartite representation of the chemical reaction network is adopted. For example, the reaction $S_1 + S_2 \rightarrow P$ is depicted as follows (see inset of the panel B); first, the chemical nodes representing S_1 and S_2 (the gray disks) are wired to the reaction node (the pink diamond), and then, an edge connects the reaction node and the chemical node of the chemical P (the gray disk). Directions of the reactions are chosen to be consistent with the steady-state flux distribution. (C to F) Four typical relaxation dynamics emerged from the minimum networks: Exponential, Metastable-Plateau, Confined-Plateau, and Power-law relaxation. The vertical and horizontal axes are on a logarithmic scale. The four dynamics are classified into three category based on the networks types that exhibit the corresponding dynamics. (G) Example distributions of the migration length (see text for definition). The top panel is the distribution over all initial conditions for one network showing metastable-plateau type relaxation, and the bottom panel is from a network showing both the plateau and power-law type relaxations. (H) Classification of the minimum networks based on the average relaxation time (see text for definition) and the average migration length.

DRAFT

$\delta = 0$ condition. On the other hand, all the type-III and type-IV networks have $\delta = 1$, i.e., in the absence of degradation no steady-state exists.

This implies that there are two types of plateau relaxation. The plateaux exhibited by type-II networks are the steady-states of the model equation without degradation (Eq. (6)). However, the plateaux exhibited by type-III and IV networks are not the steady-state of Eq. (6), but are extremely slow transient dynamics towards the final steady-state for the model *with* degradation. Indeed, we find that the type-II plateaux correspond to what we had earlier termed "metastable-plateau", while the type-III and IV plateux correspond to what we called "confined-plateau".

II. The left null-space. The rank gap does not distinguish between type-I and type-II networks (which both have $\delta = 0$). Fortunately, these networks exhibit a unique type of relaxation regardless of the initial concentrations (exponential or metastable-plateau). Thus, we can expect that the difference stems from the network structure.

Fig.2 shows two time courses of the same type-II network but starting from different initial concentrations. The plateau regions differ not only in the concentrations of the chemicals but also in the rank order of the chemical concentrations (the identical color is used for each chemical species in both figure panels).

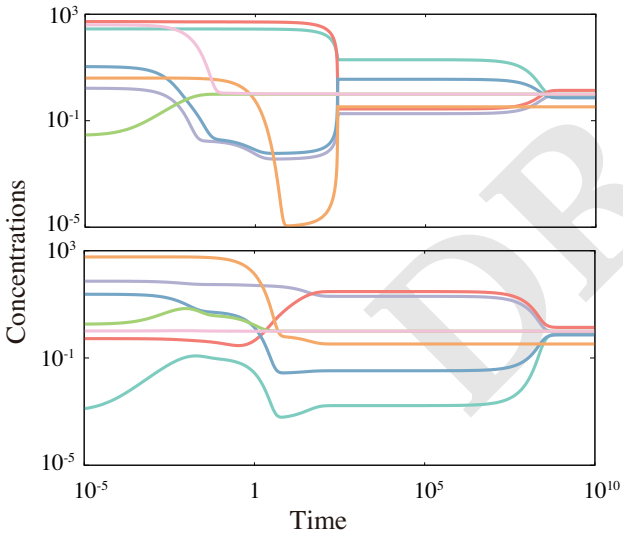

Fig. 2. Two time courses emerged from the same minimum type-II network when started with different initial concentrations. Both the vertical and horizontal axes are plotted in the logarithmic scale.

Considering that each plateau corresponds to the steady-state of $\dot{\boldsymbol{x}} = S_1\boldsymbol{v}_1 - \boldsymbol{p}$ and the model eventually reaches a unique final attractor, possible mechanisms to generate the variations are limited. We hypothesized that such type-II networks might be have combinations of chemical concentrations that, over timescales much shorter than $1/\phi$, are constrained to be almost-conserved. If different initial conditions have different values for such conserved quantities, the system might evolve towards a different steady-state of $\dot{\boldsymbol{x}} = S_1\boldsymbol{v}_1 - \boldsymbol{p}$ for each of these. We therefore computed the dimension of the left null-space of the enlarged stoichiometry matrix, $I := \dim(\mathrm{coker})\, S_2$, where $S_2 = (S_1|\boldsymbol{P})$, because the left null-space of the reaction stoichiometry represents the conserved quantities of the reaction dynamics. We found that the left null-space of all the type-I models was $\{0\}$. In contrast, all the type-II models had non-zero I[†]. Further, we found that the type-III models had no conserved quantity ($I = 0$), while the type-IV networks had at least one ($I \geq 1$).

So far, we have fully characterized the type-I and type-II networks. The type-I networks have no rank gap, $\delta = \mathrm{rank}S_1 - \mathrm{rank}S_2 = 0$, and no conserved quantity $I = \dim(\mathrm{coker}\ S_2) = 0$. Type-II networks also have no rank gap but have at least one conserved quantity, $I \geq 1$.

III. The stoichiometric cone. Now we move onto the networks with $\delta = 1$[‡]. Due to the non-zero rank gap, there is no steady-state until t becomes comparable with $1/\phi$, i.e., the spontaneous degradation starts affecting the dynamics of the concentration. Thus, in these type-III and type-IV networks, the relaxation dynamics over timescales smaller than $1/\phi$ is not to some meta-stable almost-steady-state, but a slow transient dynamics towards the final attractor with degradation. With power-law relaxation there is a slow but continuous approach towards the final attractor, while with the confined-plateau dynamics the trajectory is practically static for a long time before suddenly "jumping" to the final attractor.

We therefore hypothesized that in the power-law case, the attractor is reachable from the initial point, but the degradation term $\phi\boldsymbol{x}$ is necessary for making the final concentrations a stable steady-state. By contrast, in the confined-plateau case getting close to the attractor may itself be hindered in the absence of the degradation term. To check this, we introduce a subset of the phase space defined as

$$SC(\boldsymbol{x}_{\mathrm{ini}}) = \{\boldsymbol{x} \in \mathbb{R}_+^{|\mathscr{C}_0|} | \boldsymbol{x} = \boldsymbol{x}_{\mathrm{ini}} + S_2\boldsymbol{v}_2, \boldsymbol{v}_2 \in \mathbb{R}^{|\mathscr{R}_0|}, J^{(\mathrm{tgt})} \geq 0\}.$$

$SC(\boldsymbol{x}_{\mathrm{ini}})$ is the region in the phase space "reachable" from the initial point by controlling the chemical reaction fluxes in any way except reversely harvesting the target chemicals. Note that in the definition, no flux balance (steady-state condition) is required. Due to the non-negativity condition ($J^{(tgt)} \geq 0$) and the non-zero rank gap δ, $SC(\boldsymbol{x}_{\mathrm{ini}})$ does not coincide with the whole space $\mathbb{R}_+^{|\mathscr{C}_0|}$. This $SC(\boldsymbol{x}_{\mathrm{ini}})$ is equivalent to the stoichiometric cone in the field of chemical reaction network theory (29), and thus, we term $SC(\boldsymbol{x})$ the stoichiometric cone corresponding to the initial condition $\boldsymbol{x}$ in the following arguments.

We can now mathematically formulate our hypothesis on the relationship between reachability and relaxation behaviours: for the initial points whose stoichiometric cone $SC(\boldsymbol{x}_{\mathrm{ini}})$ contains the final attractor, relaxation becomes power-law-like, whereas, if the stoichiometric cone of the initial point does not include the attractor, plateaux appear during the relaxation (for the algorithm of judging the inclusion relationship, see Methods).

To examine the relationship between the relaxation type and the position of the initial concentrations in the phase

[†] In the present minimum reaction models, there is a lower limit of I for the models with $\mathrm{rank}S_1 = \mathrm{rank}S_2$, given by $M - N_n$, where M is the number of monomer types and N_n is the number of input chemicals . For more detail, see SI Text

[‡] Since the target formation reaction is represented as the single reaction (Eq. (5)), the rank gap between S_1 and S_2 is one at most.

DRAFT

space, we computed the average of the normalized migration length $L/L_{\max}$ separately for the initial concentrations whose stoichiometric cone does and does not contain the final attractor[§]. Fig. 3 is the scatter plot of the two average values for all the minimum networks of type-III and IV that we constructed. The migration length averaged over the initial concentrations with $\boldsymbol{x}_{\rm att} \in SC(\boldsymbol{x}_{\rm ini})$, $\langle L/L_{\max}\rangle_{\rm in}$, and with $\boldsymbol{x}_{\rm att} \notin SC(\boldsymbol{x}_{\rm ini})$, $\langle L/L_{\max}\rangle_{\rm out}$ are plotted in the horizontal and vertical axis, respectively. In all networks, $\langle L/L_{\max}\rangle_{\rm in}$ is larger than $\langle L/L_{\max}\rangle_{\rm out}$. This result supports our intuition that some trajectories are confined in a region of the phase space and cannot get closer to the attractor.

The difference between the initial conditions leading to the power-law and plateau relaxation is intuitively illustrated by a simple toy model shown in Fig.4. The model equation is given by

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} C - x^{(A)}x^{(B)} \\ (x^{(A)})^2 x^{(B)} \end{pmatrix} - \phi \boldsymbol{x}, \qquad [10]$$

where we set the direction of the reaction from left to right on Fig.4, and C is the concentration of the input chemical (the leftmost chemical in the schematic). In the model, the second reaction is the target harvesting reaction.

The first reaction produces one molecule each, but the second reaction consumes two molecules of chemical A and one molecule of chemical B. As a consequence of the imbalance, the stoichiometric cone of the model is given by [¶]

$$SC(\boldsymbol{x}_{\rm ini}) = \{\boldsymbol{x} \in \mathbb{R}^2_+ | x^{(B)} - x^{(A)} \geq x^{(B)}_{\rm ini} - x^{(A)}_{\rm ini}\}.$$

This means that the quantity $x^{(B)} - x^{(A)}$ only increase by the two reactions. In the model, we need to consume two molecules of A to consume one molecule of B. However, to increase the concentration of A, one needs to utilize the first reaction (note that the second reaction is irreversible), increasing the concentration of B by the same amount simultaneously.

Therefore, if the initial gap of the concentration is larger than the gap at the attractor, $x^{(B)}_{\rm att} - x^{(A)}_{\rm att} < x^{(B)}_{\rm ini} - x^{(A)}_{\rm ini}$, the system cannot get closer than a certain distance for $t \ll 1/\phi$. For instance, if $x^{(B)}_{\rm ini} - x^{(B)}_{\rm att} > x^{(A)}_{\rm ini}$ holds, the closest reacheable concentration of the chemical B is $x^{(B)}_{\rm ini} - x^{(A)}_{\rm ini} > x^{(B)}_{\rm att}$ [‖]. In the contrast, if the stoichiometric cone contains the attractor, $x^{(A)}_{\rm att} \ll x^{(A)}_{\rm ini}$ and $x^{(B)}_{\rm att} \gg x^{(B)}_{\rm ini}$ for example, the system can approach arbitrary close to the attractor even in the absence of the degradation term.

Typically, the quiescent dynamics on the confined plateau occur because the concentrations of one or a few chemicals become pretty low, as the chemical A in the toy example. Therefore, the concentrations at the plateau are not necessarily the closest possible state to the attractor in the stoichiometric cone but close to one of the axes of the Cartesian coordinate.

[§] Since the type-IV networks have conserved quantities, the stoichiometric cone of an initial point never contains the attractor unless the values of the conserved quantities at the initial point and the attractor are equal. This situation never happens by applying random perturbations to generate the initial points. Thus, we first search for the nearest point from the initial point, $\tilde{\boldsymbol{x}}_{\rm ini}$, that has the same values as the conserved quantities with the attractor (see Materials and Methods). Then, we checked if the attractor is in the stoichiometric cone of $\tilde{\boldsymbol{x}}_{\rm ini}$.

[¶] Note that the stoichiometic matrix S_2 for computing the stoichiometic cone in this model is

$$S_2 = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix}$$

[‖] By inversely running the first reaction.

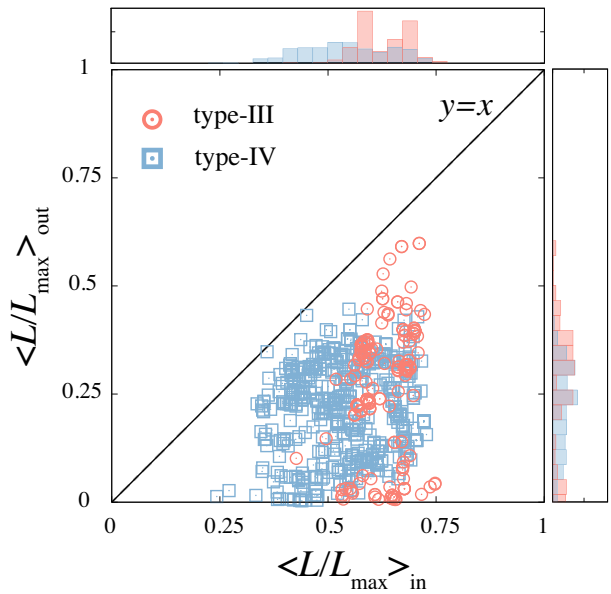

Fig. 3. The normalized migration length is averaged separately for the time courses with the initial points whose SC, defined in the text, contains (vertical axis) and does not contain (horizontal axis) the attractor. The $y = x$ line is plotted as a guide for the eye. The distributions of $\langle L/L_{\max}\rangle_{\rm in}$ and $\langle L/L_{\max}\rangle_{\rm out}$ are shown along the horizontal and the vertical axis, respectively.

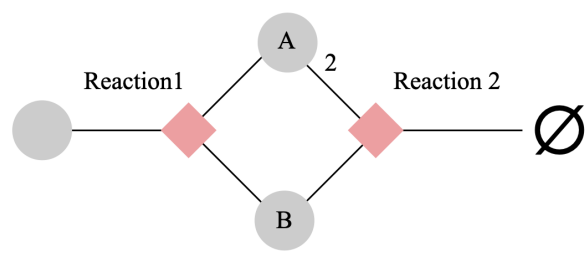

Fig. 4. A toy model to illustrate the power-law and plateau relaxation. The leftmost chemical is the external chemical whose concentration is fixed to a constant value.

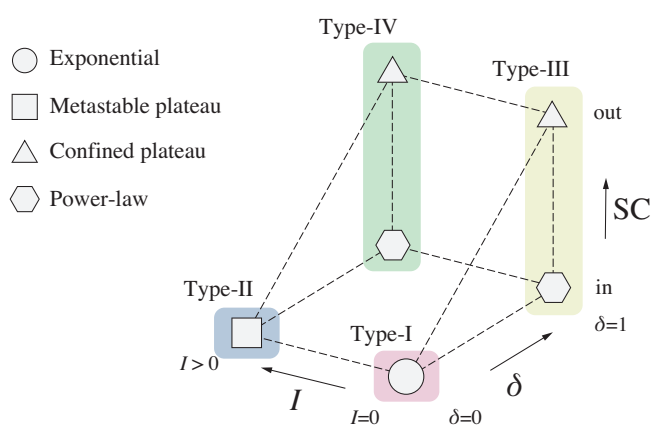

Fig. 5. The summary of the relaxation type and network-structural features. Network types are classified according to the rank gap δ and the dimension of the cokernel of the stoichiometry matrix S_2, I. In the networks with $\delta = 1$, confined plateau or power-law dynamics emerges depending on whether the final attractor is in the stoichiometric cone of the initial point.

DRAFT

A. Summary of the relaxation-structure relationship. To summarize the above classification: Networks that have zero rank gap and no conserved quantity (type-I) always exhibit exponential relaxation dynamics. Networks that have zero rank gap but some conserved quantities (type-II) always exhibit plateau relaxation. Type III and type IV networks both have a rank gap of 1, but the former have no conserved quantities, while the latter have at least one. Both these networks exhibit either power-law relaxation or plateaux, depending on the initial conditions. For both types of networks, the type of relaxation for a given initial condition is determined by whether the final attractor is reachable from that initial point. If it is, then we observe power-law relaxation, and if it is not, then we observe confined plateau relaxation. Thus, from the stoichiometric matrices associated with a given network, by computing three network-structural quantities, namely, (i) the rank gap δ, (ii) the dimension of the left null space $I = \dim(\mathrm{coker})\, S_2$, and (iii) the stoichiometric cone, we can predict whether the relaxation will be exponential, power-law or plateau-type. A graphical summary of this result is presented in Fig. 5.

Applicability of the structure-dynamics relationship for non-minimal networks

So far, we have explored the relationship between the network structure and emerging relaxation dynamics. There, we focused on the networks capable of synthesizing the target from given inputs with the fewest possible reactions. In this section we show that the three network-structural features predict the relaxation dynamics well in non-minimal networks also, and even in a real biological network.

Combining minimum networks. Here we combine two minimum networks which share the same inputs and the targets. Fig. 6 is one instance of such a combination. The two networks, labelled G and H, are minimum networks that synthesize the same target from the same input. Both networks have a non-zero rank gap δ, and in addition, the dimension of the left-null space I is zero for network G but 1 for network H. The combined network is obtained as the union of the sets of chemical species and reactions of the two networks and, thus, has redundant reactions (Fig. 6A).

Interestingly, the combined network $G \cup H$ has zero rank gap (δ=0) and no conserved quantity ($I = 0$). As a consequence, the relaxation dynamics of the combined network becomes a simple, exponential-type even though the original networks exhibit the power-law and the plateau relaxation, respectively (Fig. 6B-D).

We can also see that the effective structure of the network varies over time when we combine the two networks with different rate-constants. We construct a combined network which has rate constant unity for all reactions from network H, and rate constant $1/\tau \ll 1$ for those reactions only in G. This combined network has an effective structure with non-zero δ and I until the dynamics start to "feel" the additional part brought from G ($t \lesssim \tau$). Thus, the relaxation type becomes the power-law or the plateau type for $t \lesssim \tau$. Subsequently, once the dynamics feel the whole structure of the combined network, both δ and I are zero, and thus, the relaxation dynamics transition to the exponential-type and quickly relax to the steady-state (Fig. 7).

Large networks. Next, we ask if the three quantities, the rank gap, the left null-spaces, and the stoichiometric cone, also work as indicators of the relaxation dynamics of large networks. For this purpose, we generated a number of reaction networks consisting of a specified number of chemicals and reactions utilizing Mixed Integer Linear Programming (for details, see Methods). In the following, we deal with $|\mathscr{C}_0| = |\mathscr{R}_0| = 64$ networks constructed from the fully-connected network with $M = 2$ and $L = 8$ model. We chose the number of the input chemicals N_{in} as 2 and the number of target chemicals N_{tgt} as 1.

We classified the networks based on the rank gap (δ) and the dimension of the left null space (I) of the stoichiometric matrix. Examples of the dynamics of type-I and type-II networks are shown in Fig. S1. In line with the analysis of the minimum networks, most of the type-I and type-II networks exhibited the exponential-, and plateau-type relaxation, respectively, regardless of the initial concentrations. However, when the concentrations of the chemicals at the attractor span a wide range, several reactions having chemicals with low concentrations are considerably slowed down, and the relaxation time course cannot be described simply by the words "exponential" or "plateau". The critical values of the concentration range are studied in Fig. S2.

The relaxation dynamics get much more complicated for the networks with a non-zero rank gap $\delta > 0$. Three example time courses are shown in Fig. S3. Some trajectories are far from what we can call "power-law" or "plateau" relaxation. However, since the migration length is still capable of quantifying the differences of the relaxation dynamics, we examine the relationship between the migration length and the confinement of the trajectory in the stoichiometric cone.

Fig.8 is the scatter plot showing the relationship between $\langle L/L_{\max}\rangle_{\mathrm{in}}$ and $\langle L/L_{\max}\rangle_{\mathrm{out}}$ of each large network we constructed. In contrast to the minimum network case, the relationship $\langle L/L_{\max}\rangle_{\mathrm{in}} > \langle L/L_{\max}\rangle_{\mathrm{out}}$ is violated by a few large networks, but the broad trend remains the same – the trajectories confined in the stoichiometric cone tend to show lower values of normalized migration length.

A biological example

Lastly, we demonstrate that the features of the network structure that we have highlighted, also play a role in a real biological network, namely, the metabolic network model of glucose fermentation in *Lactococcus lactis* (Fig.9A). The model is adopted from the reference (30). The network takes up glucose (GLC) from the environment and ferments it. The target of the fermentation — either lactate (LAC), ethanol (ETOH), acetate (AC), acetoin (ACET), or butanediol (BUT) — is released into the environment. We briefly note some details of the model: the reaction rates are given by the Michaelis-Menten type functions (31), and the spontaneous degradation term ($-\phi \boldsymbol{x}$) is not implemented in the original model. The reactions depicted in the network diagram are all the reactions in the model (for a full description of the model, see the original paper).

A significant feature of the model is the balance between NAD and NADH. Suppose that a single glucose molecule is converted into either lactate or ethanol, then the production and consumption of NAD and NADH each balance. However, when a glucose molecule is converted to acetate and secreted,

DRAFT

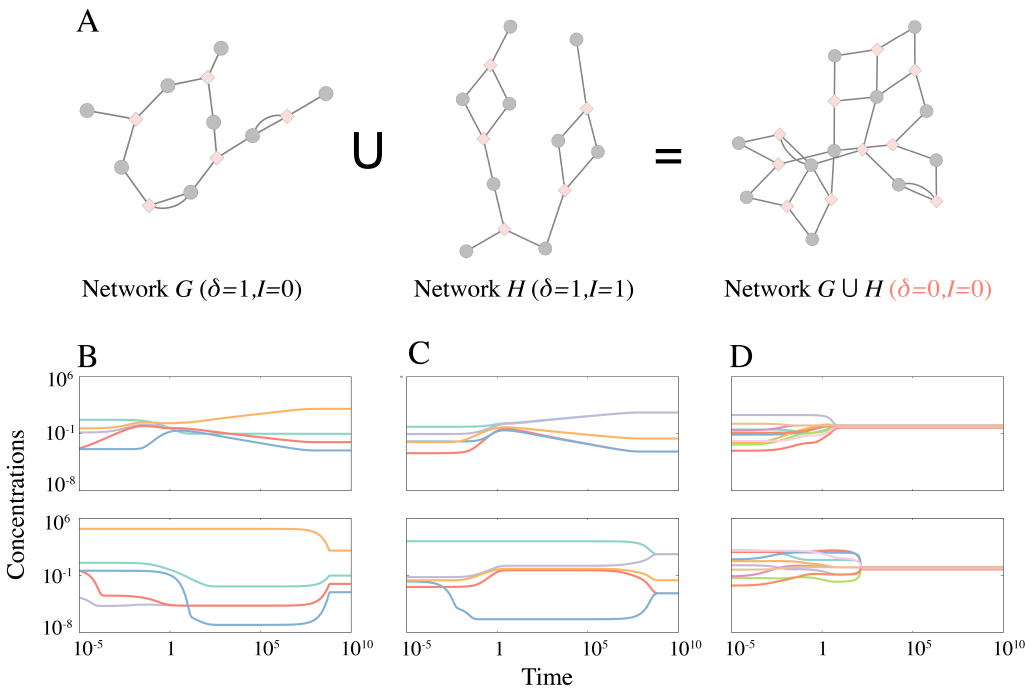

Fig. 6. A. The two networks G and H, sharing some chemical species, including the input and the target, are combined to form the union network $G \cup H$. Each network G and H has a non-zero rank gap δ, and network H has one conserved quantity. However, the combined network has zero rank gap and no conserved quantity. B and C. The time courses of the network G (B) and H (C) exhibit the power-law or the confined-plateau type relaxation due to the non-zero rank gap. D. Since δ and I are both zero in the combined network $G \cup H$, the relaxation time courses is of the exponential type.

a single NAD molecule is consumed, and a single molecule of NADH is produced. Thus, if the pathway GLC $\rightarrow$ AC is utilized, the concentrations of NAD and NADH cannot balance without the NADH oxidase (NOX) reaction (NADH $\rightarrow$ NAD) drawn separately from the main network in Fig.9A. Indeed, the stoichiometric matrix of the fermentation model will have a rank gap $\delta = 1$ when NOX is removed from the model, while with this reaction included it has a zero rank gap.

Here, we expect to observe a qualitative difference in the relaxation dynamics that originate from the rank gap by comparing the relaxation dynamics of the model with- and without the NOX reaction. However, the model without the NOX reaction can never reach a steady-state attractor. Recall that the relaxation to the attractor in the polymer models is possible because of the spontaneous degradation. We cannot introduce it into this model because the total concentrations of the following chemical pairs are set to be constant: ATP and ADP, NAD and NADH, and Coenzyme-A (CoA, which does not appear in the network diagram) and Acetyl-CoA (AcCoA). Instead, we changed the rate constant of the NOX reaction, 100 times faster and slower than the value used in (30), which results in the effective structure of the network in a certain time range to be with- and without the NOX reaction. Here, the slow down of the NOX reaction corresponds to the decrease of oxygen in the medium. Also, we set the external acetate concentration to be 10^{-4} (mM), leaving the rest of the parameter values unchanged.

Figs. 9B and C show the examples of the relaxation time course with the fast (B) and slow (C) rate constant of the NOX reaction. As seen from the figures, the relaxations are exponential-like in the model with a fast NOX reaction, and are power-law or plateau-like in the model with the slow NOX reaction.

Discussion

In the present manuscript, we have studied a class of chemical reaction networks consisting of polymer chemicals that can split and ligate with each other. The networks satisfy the law of mass conservation and can generate a variety of mass-conserving sub-networks. The (globally) minimum networks capable of synthesizing a specified set of target precursors from a set of input chemicals are computed within a reasonable time by utilizing the algorithm for solving the SAT problem.

We generated a large number of such minimum networks and simulated their chemical reaction dynamics to unveil the relationship between the network structure and the relaxation dynamics after perturbation of the steady-state. We found three types of relaxation dynamics: exponential, power-law and plateau type. In the first instance it is interesting that even though we choose all reaction rates to be unity, i.e. we do not introduce by hand multiple timescales into the system, much longer timescales much emerge spontaneously in the relaxation dynamics. In this simplified setting therefore such slow dynamics or long-term memory must arise from the network structure. We found that we could predict which relaxation behaviour will be exhibited from three structural features of the reaction networks, namely, (i) the rank gap, (ii) the dimension of the left null space, and (iii) the stoichiometric cone. We showed that these three features determine the relaxation behaviour not only in the minimum networks,

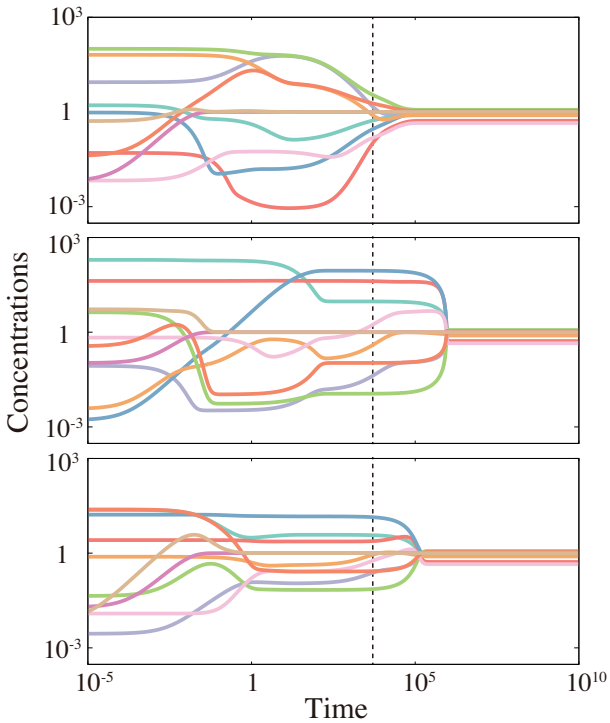

Fig. 7. Three time courses exemplifying the dynamics of the combined network $G \cup H$ with different rate-constants of the reactions (unity for reactions in H and $1/\tau$ for that only in G). The gray-dashed vertical lines indicates $t = \tau = 10^4$. The dynamics for $t \lesssim \tau$ shows the plateau- or the power-law relaxation behaviour, while after that, the state exponentially converges to the steady-state attractor, except the left bottom case.

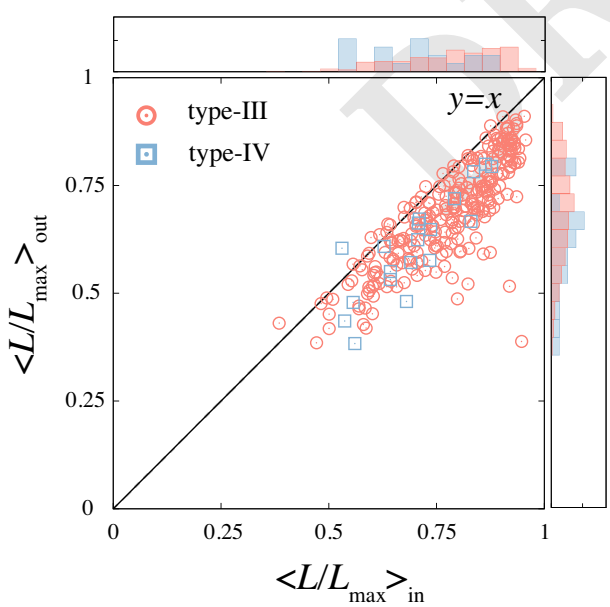

Fig. 8. The normalized migration length is averaged separately for the time courses with the initial points whose SC contains (vertical axis) and does not contain (horizontal axis) the final attractor. The $y = x$ line is plotted as a guide for the eye. The distributions of $\langle L/L_{\max}\rangle_{\rm in}$ and $\langle L/L_{\max}\rangle_{\rm out}$ are shown along the horizontal and the vertical axis, respectively.

but also in combinations of the minimum networks, in large networks with many redundant pathways, as well as in one example of a real metabolic network.

Our results therefore provide easily computable criteria for predicting whether a given chemical reaction network can lead to slow dynamics. By examining the rank gap and the left null-space, one can predict whether the concentration of the chemicals will quickly relax to a steady-state or not. A non-zero rank gap indicates that the reaction network cannot have a steady state in the intermediate timescale. except reversely harvesting the target chemicals However, as illustrated by the network combination example and the metabolic networks of *Lactococcus lactis*, the behaviours of the chemical reaction network reflect the effective network structure at the timescales of interest. Therefore, over certain timescales, metabolic, and other real chemical networks, may effectively behave as if they have a non-zero rank gap, leading to non-exponential relaxation behaviours.

Finally, we would like to emphasise the flexibility of our theoretical framework. Even with only polymer chemical species, the network size is easily controlled by manipulating the maximum length L and types of the monomers M. Further, the types of species and chemical reactions used to construct the networks can easily be expanded without affecting key chemical properties like mass conservation. For instance, one could include all species and reactions known to occur in microbes (32). The SAT formulation can be equally easily used in such a case also to construct minimum networks. We expect that our approach for constructing and analyzing such artificial reaction networks will provide a useful toolbox for exploring the extensive world of out-of-steady-state dynamics of chemical reaction networks.

Method

Generating artificial reaction networks. For the set of chemicals $\mathscr{C}$, we wish to identify the smallest set of reactions $\mathscr{R}_0 \subseteq \mathscr{R}$ such that target chemicals can be produced from the available inputs.

As input we have the set of available input chemicals $U \subset \mathscr{C}$ and the set of target chemicals $V \subset \mathscr{C}$ that must be produced. The main idea of our approach is to consider the Boolean expression

$$B = (|\mathscr{R}_0| \leq K) \wedge \bigwedge_{v \in V} \mathrm{CanMake}(v),$$

where $\mathrm{CanMake}(v)$ depends on the input chemical set U. We may then iteratively solve this SAT problem starting from $K = 1$ (to $|\mathscr{R}|$) and stop when we can satisfy the expression. This ensures that we find the global minimum solutions.

The Boolean *at most* expression, $|\mathscr{R}_0| \leq K$, can be expressed efficiently using auxiliary variables (33). What remains is the specification of the Boolean function CanMake. We can define this directly by the binary tree of reactions that leads from the inputs to the targets. For instance if the required target is AA and inputs are AB and AC, we could write

$$\begin{aligned}\mathrm{CanMake}(AA) = &(R_{AB \rightleftharpoons A,B} \wedge R_{AA \rightleftharpoons A,A}) \\ &\vee (R_{AC \rightleftharpoons A,C} \wedge R_{AA \rightleftharpoons A,A})\,.\end{aligned}$$

This fully defines a SAT-problem in the reaction R-variables. However, this approach makes exponentially long expressions for large sets of chemicals and reactions.

DRAFT

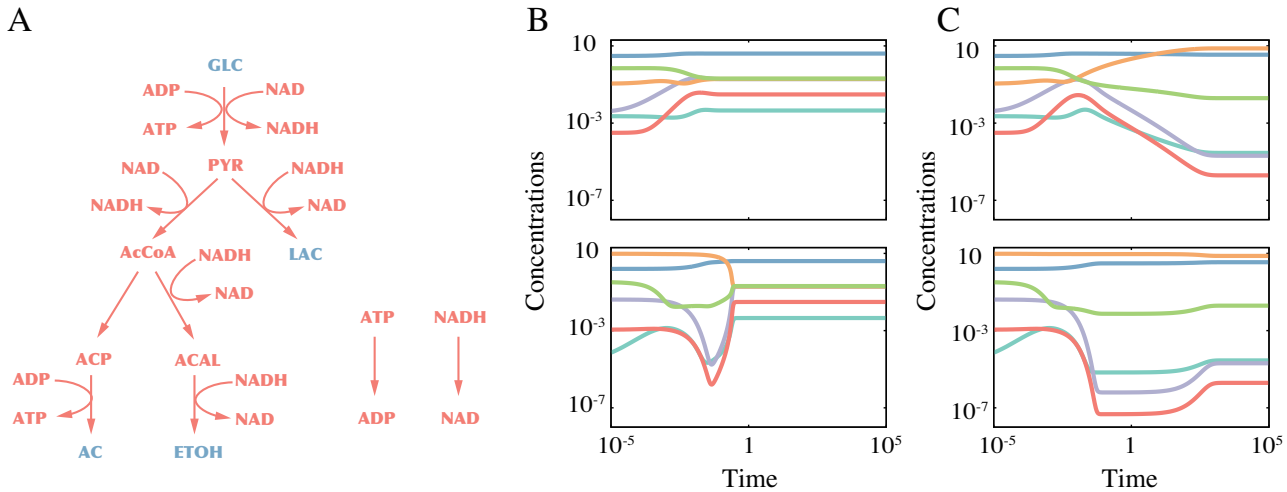

Fig. 9. (A) The metabolic model of glucose fermentation in *Lactococcus lactis* adopted from (30). The red colored metabolites and the blue colored metabolites are the intracellular- and extracellular metabolites, respectively. (B and C) The representative time courses of the model with the fast NOX reaction (B) and the slow NOX reaction (C). The time courses shown in the neighboring panels start from the same initial concentrations.

The reason for the long expressions is mainly repetitions: even in the tiny example above, the reaction variable $R_{AA\rightleftharpoons A,A}$ appeared in both sub-expressions. To remedy this, we use the concept of equisatisfiability to introduce auxiliary variables that describe if a specific chemical $c \in \mathscr{C}$ can be produced. Using these variables we can truncate the binary tree after one level. We thus intend to write the condition to make targets as

$$\bigwedge_{c\in G} c \wedge \bigwedge_{c\in\mathscr{C}} \mathrm{Relation}(c),$$

where the Boolean Relation function specifies the relationship between chemicals as specified by the set of reactions. For instance, for the above example:

$$\mathrm{Relation}(AA) = AA \odot \left[A \wedge R_{AA\rightleftharpoons A,A}\right], \qquad [11]$$

i.e. to make AA you need available A and the reaction $AA \rightleftharpoons A, A$. Here, $\odot$ denotes *exclusive nor* indicating for $X \odot Y$ that either both X and Y are TRUE or both are FALSE. Likewise, for example

$$\begin{aligned}\mathrm{Relation}(A) &= A \odot \big[(AA \wedge R_{AA\rightleftharpoons A,A}) \vee (AB \wedge R_{AB\rightleftharpoons A,B}) \\ &\qquad \vee (AC \wedge R_{AC\rightleftharpoons A,C})\big] \\ &= A \odot \big[(AA \wedge R_{AA\rightleftharpoons A,A}) \vee R_{AB\rightleftharpoons A,B} \vee R_{AC\rightleftharpoons A,C}\big],\end{aligned}$$

where we used the fact that AB and AC are input chemicals.

With this approach, however, we need to restrict the direction of the reactions: reactions should only be used in one direction as the network of reactions from inputs to targets should be *directed* and *acyclic* (a DAG). To fix the directionality, we introduce Boolean direction variables $\{D_i\}$ with the convention that $D = \mathrm{TRUE}$ indicates breakdown reactions and $D = \mathrm{FALSE}$ indicates assembly. If we denote the reaction $AA \rightleftharpoons A, A$ by k, Eq. (11) thus becomes $\mathrm{DRelation}(AA) = AA \odot [A \wedge R_k \wedge \neg D_k]$, whereas the DRelation for A would use D_k without the negation. When there is no reaction, we do not care about the direction of the (non-existent) reaction and we thus include a term that enforce $D_k = \mathrm{TRUE}$ if there is no reaction, i.e. $\neg R_k \Rightarrow D_k = R_k \vee D_k$.

The final reformulated expression is

$$B = (|\mathscr{R}_0| \le K) \wedge \bigwedge_{c\in G} c \wedge \bigwedge_{c\in\mathscr{C}} \mathrm{DRelation}(c) \wedge \bigwedge_{i\in\mathscr{R}} (R_i \vee D_i).$$

The only remaining complexity is the requirement that the solution of reactions is acyclic. We implement this requirement outside the language of Boolean expressions and simply discard such solutions. If only cyclic solutions are found at a certain K, these solutions are disallowed at larger K, thus avoiding a large number of degenerate cyclic solutions at large K.

All solutions to the SAT-problem are found using the GLUCOSAMINE solver via the BOOLEXPR interface.

Checking if the attractor is in the stoichiometric cone. For checking if the attractor is in the stoichiometric cone of the initial point, $SC(\boldsymbol{x}_{\mathrm{ini}})$, we computed the L^∞ norm between the attractor and the points in the stoichiometric cone and minimize the norm, by utilizing the following linear optimization problem.

$$\begin{aligned}\text{minimize} \quad & \max\left|x^{(i)}_{\mathrm{att}} - (x^{(i)}_{\mathrm{att}} + \sum_j (S_2)_{ij} v_2^{(j)})\right|, (0 \le i < |\mathscr{C}_0|) \\ \text{subject to} \quad & x^{(i)}_{\mathrm{att}} + \sum_j (S_2)_{ij} v_2^{(j)} > 0, (0 \le i < |\mathscr{C}_0|) \\ & J^{(\boldsymbol{p}r)} > 0\end{aligned}$$

If the minimum distance is smaller than 10^{-12}, we regard that the attractor is in the stoichiometric cone and *vice versa*.

Finding $\tilde{\boldsymbol{x}}_{\mathrm{ini}}$. For obtaining $\tilde{\boldsymbol{x}}_{\mathrm{ini}}$, the nearest point to the initial point with the same value of the conserved quantities $\boldsymbol{Q}$ with that of the attractor, we utilized the following quadratic optimization problem with $\boldsymbol{\delta x}$ as variables

$$\begin{aligned}\text{minimize} \quad & \|\boldsymbol{\delta x}\|^2 \\ \text{subject to} \quad & \boldsymbol{Q}(\boldsymbol{x}_{\mathrm{ini}} + \boldsymbol{\delta x}) = \boldsymbol{Q}(\boldsymbol{x}_{\mathrm{att}}) \\ & \boldsymbol{x} \ge \boldsymbol{0}.\end{aligned}$$

and obtained $\tilde{\boldsymbol{x}}_{\mathrm{ini}}$ as $\tilde{\boldsymbol{x}}_{\mathrm{ini}} = \boldsymbol{x}_{\mathrm{ini}} + \boldsymbol{\delta x}$.

DRAFT

Generating networks with a specified size. Due to the computational cost, we adopt a mixed-integer linear (MIL) optimization approach for generating the large (non-minimal) networks, instead of the SAT approach. To obtain a network with $|\mathscr{C}_0|$ chemicals, $|\mathscr{R}_0|$ reactions, $N_{\rm in}$ inputs, and $N_{\rm tgt}$ target precursors by solving the following MIL problem:

$$
\begin{aligned}
\text{minimize} \quad & \mathbf{1}_{|\mathscr{C}|} \cdot \boldsymbol{\xi} \\
\text{subject to} \quad & \tilde{S}(\boldsymbol{v}_s^{(+)} - \boldsymbol{v}_s^{(-)}) + U_{\max}\boldsymbol{n}_s + \boldsymbol{\xi} \geq \boldsymbol{g}, \\
& 0 \leq \boldsymbol{v}_s^{(\pm)} \leq M\boldsymbol{\sigma}, \\
& 2 - \sum_s \boldsymbol{n}_s - \boldsymbol{g} - \tilde{T}\boldsymbol{\sigma} \leq M(\mathbf{1}_{|\mathscr{C}|} - \boldsymbol{\theta}) \\
& 3\sigma \leq |\tilde{S}|\boldsymbol{\theta} \\
& \mathbf{1}_{|\mathscr{C}|} \cdot \boldsymbol{n}_s = 1 \\
& \mathbf{1}_{|\mathscr{C}|} \cdot \boldsymbol{n}_p = N_{\rm tgt} \\
& \mathbf{1}_{|\mathscr{C}|} \cdot \boldsymbol{\theta} = |\mathscr{C}_0| \\
& \mathbf{1}_{|\mathscr{R}|} \cdot \boldsymbol{\sigma} = |\mathscr{R}_0|
\end{aligned}
\qquad [12]
$$

where the continuous variables are $\boldsymbol{\xi} \geq 0$ and $\boldsymbol{v}_s^{(\pm)} \geq 0$, and the bool variables are $\boldsymbol{n}_s, \boldsymbol{n}_p, \boldsymbol{\theta}, \boldsymbol{\sigma}$. $\boldsymbol{\xi}$ and $\boldsymbol{v}_s^{(\pm)}$ are the slack variable and the forward- and backward reaction flux vectors under the environment s, respectively. $\boldsymbol{n}_s$ and $\boldsymbol{n}_p$ denotes the choice of the input and the target. If the ith element of the vector $\boldsymbol{n}_s$ ($\boldsymbol{n}_p$) is one, the ith chemical is the input supplied under the environment s (the target). $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$ denotes which chemicals ($\boldsymbol{\theta}$) and reactions ($\boldsymbol{\sigma}$) are used. If the entity is unity, it is used.

$\tilde{S}$ and $U_{\max}$ are the stoichiometric matrix of all reactions in $\mathscr{R}$, and the maximum uptake rate of the input chemicals. $\tilde{T}$ and $|\tilde{S}|$ are defined as $\tilde{T}_{ij} = |\mathrm{sgn}(\tilde{S}_{ij})|$ and $|\tilde{S}|_{ij} = |\tilde{S}_{ij}|$. $\mathbf{1}_n$ is the abbreviated representation of n-dimensional vector with all the elements being unity.

Each constraint of Eq. (12) represents the following conditions. (First line) The total production rate of the ith chemical should be larger than p_i ($= 1$ if it is one of the targets and $= 0$ otherwise). (Second line) The rth reaction flux can be turned-on if σ_r is one. (Third line) If θ_i can be unity only if two or more (one or more for the inputs and the targets) reactions connected to the ith chemical are turned-on. (Forth line) To turn the r th chemical on, the inputs and the target of the reaction should be in the network. The fifth to the eighth line set the number of the inputs, targets, total chemicals, and total reactions.

Eq. (12) is always feasible and if the optimal value of the objective is smaller than 10^{-12}, the original network $\{\mathscr{C}, \mathscr{R}\}$ has a subnetwork of the specified size. For generating multiple networks, we introduce an additional constraint

$$\sigma_i = 0, \ i \in K_{\rm off}$$

where the index set $K_{\rm off}$ consists of randomly-chosen indices.

Acknowledgments

The authors thank Chikara Furusawa for fruitful discussion. This work is supported by the research grant (00028054) from VILLUM FONDEN and JSPS KAKENHI Grant Number JP 21K20626 and JP 22K15069. J.B.K. acknowledges funding from the Novo Nordisk Foundation, Grant Agreement NNF20OC0062047. N.M. acknowledges funding from Novo Nordisk foundation, Grant Agreement NNF21OC0068775. S.K. acknowledges funding from the Department of Atomic Energy, Government of India (Project Identification No. RTI 4006) and the Simons Foundation (Grant No. 287975).

1. P Érdi, J Tóth, *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models.* (Manchester University Press), (1989).
2. HB Callen, *Thermodynamics and an Introduction to Thermostatistics.* (American Association of Physics Teachers), (1998).
3. KJ Laidler, J Keith, , et al., *Chemical kinetics.* (McGraw-Hill New York) Vol. 2, (1965).
4. NG Van Kampen, *Stochastic processes in physics and chemistry.* (Elsevier) Vol. 1, (1992).
5. A Varma, BO Palsson, Metabolic capabilities of escherichia coli: I. synthesis of biosynthetic precursors and cofactors. *J. theoretical biology* **165**, 477–502 (1993).
6. A Varma, BO Palsson, Metabolic capabilities of escherichia coli ii. optimal growth patterns. *J. Theor. Biol.* **165**, 503–522 (1993).
7. JS Edwards, BO Palsson, Robustness analysis of the escherichiacoli metabolic network. *Biotechnol. progress* **16**, 927–939 (2000).
8. JS Edwards, RU Ibarra, BO Palsson, In silico predictions of escherichia coli metabolic capabilities are consistent with experimental data. *Nat. biotechnology* **19**, 125–130 (2001).
9. MW Covert, CH Schilling, B Palsson, Regulation of gene expression in flux balance models of metabolism. *J. theoretical biology* **213**, 73–88 (2001).
10. DW Erickson, et al., A global resource allocation strategy governs growth transition kinetics of escherichia coli. *Nature* **551**, 119–123 (2017).
11. M Scott, CW Gunderson, EM Mateescu, Z Zhang, T Hwa, Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**, 1099–1102 (2010).
12. S Klumpp, Z Zhang, T Hwa, Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**, 1366–1375 (2009).
13. Y Himeoka, K Kaneko, Theory for transitions between exponential and stationary phases: universal laws for lag time. *Phys. Rev. X* **7**, 021049 (2017).
14. Y Himeoka, B Gummesson, MA Sørensen, SL Svenningsen, N Mitarai, Distinct survival, growth lag, and rrna degradation kinetics during long-term starvation for carbon or phosphate. *mSphere* pp. e01006–21 (2022).
15. O Gefen, O Fridman, I Ronin, NQ Balaban, Direct observation of single stationary-phase bacteria reveals a surprisingly long period of constant protein production activity. *Proc. Natl. Acad. Sci.* **111**, 556–561 (2014).
16. Y Kaplan, et al., Observation of universal ageing dynamics in antibiotic persistence. *Nature* **600**, 290–294 (2021).
17. I Levin-Reisman, et al., Automated imaging with scanlag reveals previously undetectable bacterial growth phenotypes. *Nat. Methods* **7**, 737 (2010).
18. E Biselli, SJ Schink, U Gerland, Slower growth of escherichia coli leads to longer survival in carbon starvation due to a decrease in the maintenance rate. *Mol. Syst. Biol.* **16**, e9478 (2020).
19. K Kaneko, Recursiveness, switching, and fluctuations in a replicating catalytic network. *Phys. Rev. E* **68**, 031909 (2003).
20. D Segré, D Ben-Eli, D Lancet, Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. *Proc. Natl. Acad. Sci.* **97**, 4112–4117 (2000).
21. H Jeong, B Tombor, R Albert, ZN Oltvai, AL Barabási, The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
22. E Ravasz, AL Barabási, Hierarchical organization in complex networks. *Phys. review E* **67**, 026112 (2003).
23. P Schwaller, et al., Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS central science* **5**, 1572–1583 (2019).
24. J Nam, J Kim, Linking the neural machine translation and the prediction of organic chemistry reactions. *arXiv preprint arXiv:1612.09529* (2016).
25. M Segler, M Preuß, MP Waller, Towards" alphachem": Chemical synthesis planning with tree search and deep neural network policies. *arXiv preprint arXiv:1702.00020* (2017).
26. J Zaldivar, J Nielsen, L Olsson, Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Appl. microbiology biotechnology* **56**, 17–34 (2001).
27. B Hahn-Hägerdal, M Galbe, MF Gorwa-Grauslund, G Lidén, G Zacchi, Bio-ethanol–the fuel of tomorrow from the residues of today. *Trends biotechnology* **24**, 549–556 (2006).
28. C Moore, S Mertens, *The nature of computation.* (OUP Oxford), (2011).
29. M Feinberg, *Foundations of chemical reaction network theory.* (Springer), (2019).
30. MH Hoefnagel, et al., Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysisthe genbank accession number for the sequence reported in this paper is ay046926. *Microbiology* **148**, 1003–1013 (2002).
31. A Cornish-Bowden, *Fundamentals of enzyme kinetics.* (John Wiley & Sons), (2013).
32. M Kanehisa, M Furumichi, Y Sato, M Ishiguro-Watanabe, M Tanabe, Kegg: integrating viruses and cellular organisms. *Nucleic acids research* **49**, D545–D551 (2021).
33. C Sinz, *International conference on principles and practice of constraint programming.* (Springer), pp. 827–831 (2005).

DRAFT