

oooo
oooo
oooooooo
oo
ooo

ooooo

Complex Structures and Dynamics

Applications to Chemistry, Biology, Social Science and Economics

Markus Kirkilionis

July 19, 2012

oooo
oooo
oooo
oooooooooooo
oo
oo
ooo

ooooo

Abstract

We introduce rules which describe how events are updating the state spaces introduced. The rules are either with respect to a local structure where neighborhood (not necessarily in 'physical space') plays a role, or the rules apply equally to every object in the system. The latter is equivalent to rules defined on the complete graph. In the chemical literature rules applied to the complete graph are called 'reactions'. Reaction kinetics serves us as a blueprint for rules which are independent of a connectivity structure and therefore need not keep track of the location of an entity.

○○○○
○○○○
○○○○○○○○○○
○○
○○○

○○○○○

Introduction

- Rule-based descriptions

- Analysis of the opinion formation model

- Basic definitions related to reaction networks

- A simple example

- The reaction graph, and the complex-species graph

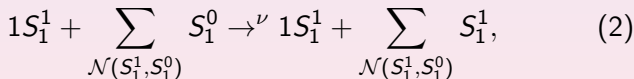
The Chemical Master Equation

- Basic definition

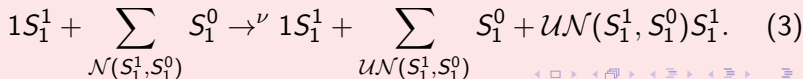


Opinion formation

We assume there are 2 possible transformations of n many individuals having 2 states only. The 'stoichiometric coefficient' 1 is interpreted as 'anyone with the same probability, but the same'. The neighborhood operator $\mathcal{N}(S_1^1, S_1^0)$ takes into account all 0-state neighbours of S_1^1 . $\mathcal{UN}(S_1^1, S_1^0)$ choses only one 0-state neighbour with a uniform probability distribution.



or the second rule better replaced by

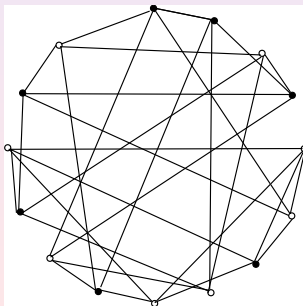


●●○○
○○○○
○○○○○○○○○○
○○
○○
○○

○○○○○

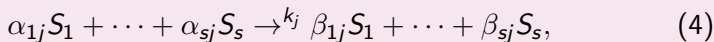
Individual Interaction Graph

In the molecular interaction graph each individual is represented as a node. Each individual has two states indicated by black and white. The number of nodes is fixed. In this example the degree distribution is such that every individual has exactly 4 links.





We assume there are r many possible transformations of molecules into each other, including also the possibility that molecules leave or enter the system. We also assume that only integer combinations of molecular species can be formed. In this case we can write each event/reaction r_j in the form

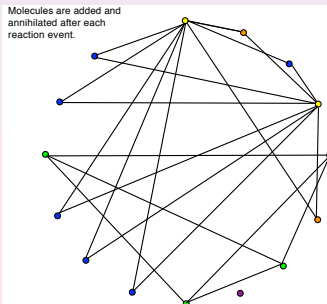


with $j = 1, \dots, r$.



Molecular Interaction Graph

In the molecular interaction graph each molecule is represented as a node. There are different types of molecules which are color coded. If each type could react with each type the molecular interaction graph would always be the complete graph.



○○○○
●○○○
○○○○○○○○○○
○○
○○

○○○○○

Pair approximation

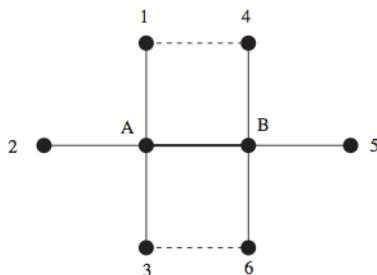
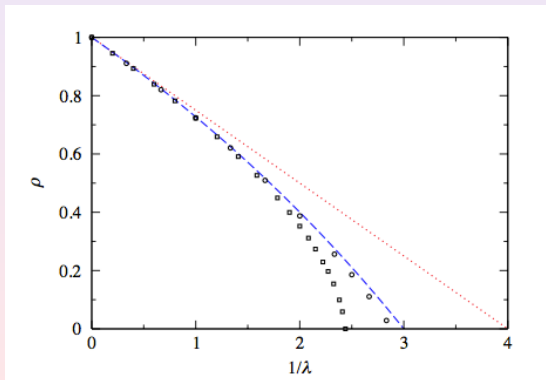


Fig. 2. An arbitrarily chosen link and its nearest neighbourhood within a homogeneous network characterized by the degree distribution $P(k) = \delta_{k4}$. The dashed lines indicate the connections which are present in the case of a square lattice.



Equilibrium analysis

Shown are equilibrium states for different values of $\lambda = \frac{\nu}{\delta}$ and their approximations.



```

○○○
○○●○
○○○○○○○○○○
○○
○○
○○○

```

```

○○○○○

```

Cluster approximation

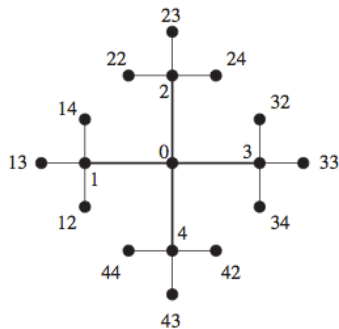
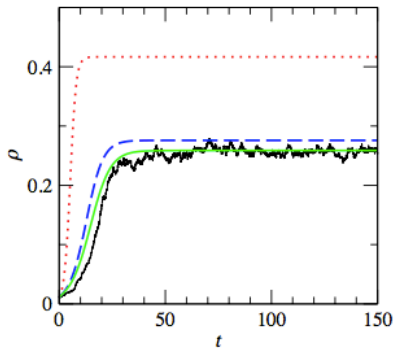
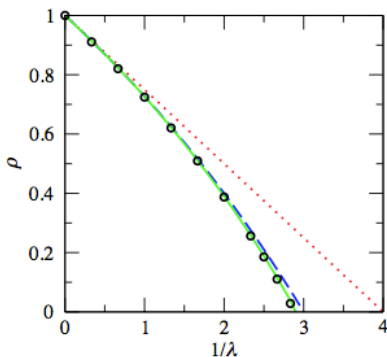


Fig. 4. An arbitrary node (denoted by 0) with its corresponding star-like fundamental cluster within a homogeneous random network of degree $K = 4$.

○○○○
 ○○○●
 ○○○○○○○○
 ○○
 ○○
 ○○

○○○○○

Random Network Simulation and Approximation





Definition

[Formal Sum]

Let I be an arbitrary finite index set. We denote by \mathbb{Z}_+^I the set of all formal sums with nonnegative integers, i.e.

$$\mathbb{Z}_+^I := \left\{ \sigma = \sum_{i \in I} \sigma_i i \mid \sigma_i \text{ is a nonnegative integer} \right\}.$$



The modeling idea of the formal sum is to count with σ_i the number of times objects in the finite set I are occurring in some system or are involved in some system during some event. In case of reaction networks the set of reacting species (always denoted by \mathcal{S}) is the natural choice for I as the number of such species is finite for all practical purposes. Usually the number of molecules that are reacting should be countable, motivating the choice of integers to define the formal sum. There is however frequently another choice made in the literature, replacing \mathbb{Z}_+^I by \mathbb{R}_+^I which is the set of formal sums constructed with the nonnegative real numbers, i.e. $\sigma_i \in \mathbb{R}_+$, see Definition 2.



Definition

[Chemical Reaction Network]

A *chemical reaction network* is a triple $(\mathcal{S}, \mathcal{C}, \mathcal{R})$, where \mathcal{S} is the set of chemical species, $\mathcal{C} \subseteq \mathbb{Z}_+^S$ is the set of complexes, and \mathcal{R} contains ordered binary relations on \mathcal{C} , each denoted $P \rightarrow P'$, representing the set of reactions in the network. Here \mathcal{P} is called the *source complex* and \mathcal{P}' is called the *target complex* of the reaction $P \rightarrow P'$. Moreover, the set \mathcal{R} must satisfy the following three conditions:

1. It cannot contain elements of the form $P \rightarrow P$.
2. For any $\mathcal{P} \in \mathcal{C}$ there exists some $P' \in \mathcal{C}$ such that either $P \rightarrow P'$ or $P' \rightarrow P$.
3. The union of the supports of all $P \in \mathcal{C}$ is \mathcal{S} .



It is useful to fix the dimensions of the different sets occurring in Definition 3. These conventions will be valid for the whole chapter. First the set of species \mathcal{S} contains $s \in \mathbb{N}$ elements, i.e. $|\mathcal{S}| = s$. We can refer to single species by writing $\mathcal{S} = \{S_1, \dots, S_s\}$. We assume there are r many possible transformations of molecules into each other, including also the possibility that molecules leave or enter the system. This means $|\mathcal{R}| = r$. Note that Definition 3 assumes that only integer combinations of molecular species are formed. In this case we can write each event/reaction r_j with the help of two formal sums as

$$\alpha_{1j}S_1 + \dots + \alpha_{sj}S_s \rightarrow \beta_{1j}S_1 + \dots + \beta_{sj}S_s, \quad (5)$$

with $j = 1, \dots, r$. The integer coefficients α_{ij} and β_{ij} represent the number of S_i molecules participating in j -th reaction at reactant and product stages, respectively. They are called *stoichiometric coefficients* (of the source and target complex, respectively).



Referring to Definition 3 we have introduced the notation $P = \alpha_{1j}S_1 + \cdots + \alpha_{sj}S_s$ and $P' = \beta_{1j}S_1 + \cdots + \beta_{sj}S_s$ for some $P, P' \in \mathcal{C}$ being in a source-target relationship defined by the j -th reaction. Instead of using formal sums, an object introduced in the chemical literature, we can equivalently associate complexes with column vectors of length s . We write

$$P = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{sj} \end{pmatrix}, \quad P' = \begin{pmatrix} \beta_{1j} \\ \vdots \\ \beta_{sj} \end{pmatrix}.$$



Both the formal sum and the vector notation for complexes will be used interchangeably. We can also label and count the elements in \mathcal{C} . If we set $|\mathcal{C}| = c$ then c is less or equal to $2r$, depending on whether some source and target complexes are identical in different reactions. With this convention we write

$$\mathcal{C} = \{C_1, \dots, C_c\}.$$

Note in this notation we do not differentiate between the reactant (source) and product (target) aspect (the α and β notation for reactions above, respectively), but only look inside which complexes these species dependent integer coefficients appear. We do however fix in a moment an ordering of the complex set after fixing an order of the set of reactions, see below. The stoichiometric coefficients building up all complexes are then assembled inside a $(s \times c)$ -matrix \mathcal{V} . We call $\mathcal{V} = (C_1, \dots, C_c)$ the *complex matrix*.



Using the column vector notation we should be able to rewrite the definition of the stoichiometric matrix \mathcal{N} as well. There is however a notational difficulty in case $c < 2r$, i.e. if some complexes of the system $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ are identical in different reactions. To solve this problem we fix an ordering of complex indices in the following way: First we fix a reaction set ordering, running from $i = 1$ to $i = r$. Next C_1 is always the source complex of reaction 1, C_2 the target complex of reaction 1 (note that because ' $P \rightarrow P$ ' is not allowed indeed $C_1 \neq C_2$). We then increase index i by one while running through all reactions in source to target direction and encountering a complex that is not equal to a previous one already having a label. Let us introduce the notation where we introduce indices $j_s(i)$ and $j_t(i)$, with i running from $i = 1$ to $i = r$, so running through all reactions.



Moreover $j_s(i)$ is defined as returning the index of the source, and $j_t(i)$ the index of the target complex of reaction i . With this notation we define

$$\mathcal{N} := (C_{j_t(1)} - C_{j_s(1)}, \dots, C_{j_t(r)} - C_{j_s(r)}).$$

The matrix \mathcal{N} is called the stoichiometric matrix. It defines one of the fundamental relationships of reaction networks.



It is convenient to introduce the *stoichiometric subspace* T of the set of reactions \mathcal{R} , closely linked to the stoichiometric matrix \mathcal{N} :

$$\begin{aligned} T &:= \text{span}\{P' - P \mid P \rightarrow P' \in \mathcal{R}\} \\ &= \text{span}\{C_{j_t(1)} - C_{j_s(1)}, \dots, C_{j_t(r)} - C_{j_s(r)}\}. \end{aligned}$$

After introducing dynamics on the reaction system $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ later, for example by introducing reaction events, we will see that the stoichiometric subspace $T \subset \mathbb{R}^s$ is the only reachable part of the state space, i.e. all possible species numbers, and after a continuum limit, also species concentrations are confined to T .



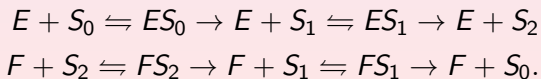
It is convenient to introduce the *stoichiometric subspace* T of the set of reactions \mathcal{R} , closely linked to the stoichiometric matrix \mathcal{N} :

$$\begin{aligned} T &:= \text{span}\{P' - P \mid P \rightarrow P' \in \mathcal{R}\} \\ &= \text{span}\{C_{j_t(1)} - C_{j_s(1)}, \dots, C_{j_t(r)} - C_{j_s(r)}\}. \end{aligned}$$

After introducing dynamics on the reaction system $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ later, for example by introducing reaction events, we will see that the stoichiometric subspace $T \subset \mathbb{R}^s$ is the only reachable part of the state space, i.e. all possible species numbers, and after a continuum limit, also species concentrations are confined to T .



An illustrative example we consider throughout this chapter is a well-known model describing the activity of the mitogen-activated protein kinase (MAPK). MAPKs play an important part in the signalling processes of eukaryotic cells by intervening with a multitude of proteins and phosphorylating them. They themselves undergo phosphorylation by a MAPK/ERK kinase (MEK) and dephosphorylation by a phosphatase. In this model we use a notation for species adapted to the biochemical interpretation. Let S_i , with $i = 0$, be the MAPK kinase, and with subscripts $i = 1$ and $i = 2$ the molecules with single and double phosphorylation, respectively. The symbols E and F represent MEK and the phosphatase. The reaction scheme for the model is:





Therefore we have with the obvious notational identifications:

$$\mathcal{S} = \{S_1, \dots, S_9\} = \{E, F, S_0, S_1, S_2, ES_0, ES_1, FS_1, FS_2\},$$

i.e. $s = 9$ in this example. Moreover, using the formal sum notation, we have

$$\begin{aligned} \mathcal{C} &= \{C_1, \dots, C_{10}\} \\ &= \{E + S_0, ES_0, E + S_1, ES_1, E + S_2, F + S_2, FS_2, F + S_1, FS_1, F + S_0\}, \end{aligned}$$

i.e. $c = 10$ in this example, and all stoichiometric coefficients are 1 and therefore do not appear. There are 12 reactions, i.e. $r = 12$.

Note that we used the symbol " \rightleftharpoons " to denote a *reversible reaction*. This means there are two reactions in opposite direction of each other, the complexes left and right of these arrows can be simultaneously target and source complexes.



We are now able to introduce a first graph associated with reaction networks, the directed *reaction graph* $\vec{G}_R = (V, \vec{E})$. We simply define the vertex set $V_{\vec{G}_R}$ as the set of complexes of the reaction network, i.e. $V(\vec{G}_R) = \mathcal{C}$. There is a directed edge between two complexes whenever there is a reaction, pointing from the source complex to the target complex. For our MAPK example reaction system this means

$$\vec{E}(\vec{G}_R) = \{(C_1, C_2), (C_2, C_1), (C_2, C_3), (C_3, C_4), (C_4, C_3), (C_4, C_5), (C_6, C_7), (C_7, C_6), (C_7, C_8), (C_8, C_9), (C_9, C_8), (C_9, C_{10})\}.$$

○○○○
○○○○
○○○○○○○○○○
○○
○○
●●○

○○○○○

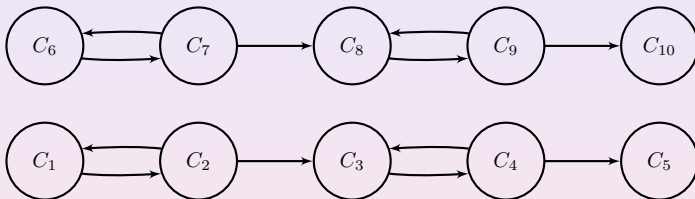


Figure: The directed reaction graph \vec{G}_R of the MAPK reaction network example.

```

○○○○
○○○○
○○○○○○○○○○
○○
○○
○○●

```

```

○○○○○

```

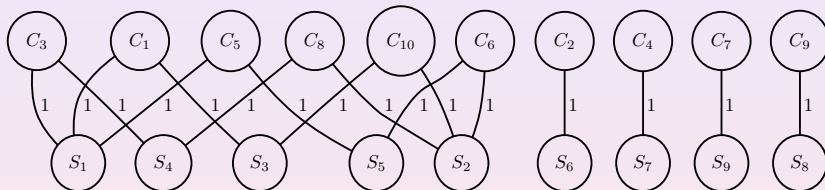


Figure: The undirected bipartite complex-species graph G_{CS} of the MAPK reaction network example.



A natural way to describe the changing number of molecules of each species¹ S_i is by introducing a continuous-time Markov jump process for each reaction. Let $N_i(t)$ be a random variable that represents the number of molecules of species S_i at time t , and let N denote the vector of N_i s. Further, let $P(t, n)$ be the joint probability that $N(t) = n$, i.e., $N_1 = n_1, N_2 = n_2, \dots, N_s = n_s$. Clearly the state of the system at any time is now a point in \mathbb{Z}_0^s , where \mathbb{Z}_0 is the set of non-negative integers.

¹That is to assign a law of evolution to the system's state which components are attached to each node and representing a species' number occurring in the reaction volume.



Formally the master equation that governs the evolution of P is then

$$\frac{d}{dt}P(t, n) = \sum_{m \in S(n)} \mathcal{R}(m, n) \cdot P(t, m) - \sum_{m \in T(n)} \mathcal{R}(n, m) \cdot P(t, n), \quad (6)$$

where $\mathcal{R}(m, n)$ is the probability per unit time of a transition from state m to state n , $\mathcal{R}(n, m)$ is the probability per unit time of a transition from state n to state m , $S(n)$ is the set of all states that can terminate at n after one reaction step, and $T(n)$ is the set of all states reachable from n in one step of the feasible reactions.

The notation is meant to suggest the source and target states at n ; one could also call $S(n)$ the predecessors of state n and $T(n)$ the successors of state n . The predecessor states must be non-negative for production reactions and positive for conversion, degradation and catalytic reactions.



Similar bounds on the target states are naturally enforced by zero rates of reaction when the reactants are absent. The sets $S(n)$ and $T(n)$ are easily determined using the reaction graph G_R . Let again $\mathcal{C} = \{C_1, \dots, C_c\}$ be the set of complexes, i.e. the nodes of G_R . Let $r_l \in E(G_R)$ (the edge set of G_R), $1 \leq l \leq r$, whenever there exist two complexes with a reaction defined between them. Define the incidence matrix² \mathcal{I} by

$$\mathcal{I}_{il} = \begin{cases} +1 & \text{if } r_l \text{ is incident at } C_i \text{ and is directed toward it,} \\ -1 & \text{if } r_l \text{ is incident at } C_i \text{ and is directed away from it,} \\ 0 & \text{otherwise.} \end{cases}$$

²This is a matrix with c rows (number of complexes) and r columns (number of reactions).

○○○
○○○
○○○○○○○○
○○
○○○

○○●○

The sets $S(n)$ and $T(n)$ can now be determined using the G_R graph structure. It follows from the definition of \mathcal{V} and \mathcal{I} that the l th reaction r_l between say $C_i \rightarrow C_j$ induces a change $\Delta n^{(l)} = \mathcal{V}\mathcal{I}_{(l)}$ in the number of molecules of all species after one reaction event, where the subscript denotes the l th column of \mathcal{I} . Therefore the state $m = n - \mathcal{V}\mathcal{I}_{(l)}$ is a source or predecessor to n under one step of the l th reaction. Similarly, states of the form $m = n + \mathcal{V}\mathcal{I}_{(l)}$ are reachable from n in one step of the l th reaction.



With this insight we can now sum over reactions instead of sources and targets to get

$$\frac{d}{dt}P(t, n) = \sum_{l=1}^r \mathcal{R}_l(n - \mathcal{V}\mathcal{I}_{(l)}) \cdot P(t, n - \mathcal{V}\mathcal{I}_{(l)}, t) - \sum_{l=1}^r \mathcal{R}_l(n) \cdot P(t, n), \quad (7)$$

with $\mathcal{R}_l(n)$ now being the probability per unit of time that a reaction event r_l is taking place and the numbers of the different species in \mathcal{S} before the event is given by the vector n . The master equation is now conforming to the complex structure and describes the reaction network on the microscopic level. It can be used to investigate how noise is occurring in the system when particle numbers are small, and how this noise depends on the network structure. We can use the master equation for direct simulation of the reaction network.