

Cooperative Assembly Systems*

Vincent Danos^{1**}, Heinz Koepl², and John Wilson-Kanamori¹

¹ LFCS, School of Informatics, University of Edinburgh

² ETH Zurich

Abstract. Several molecular systems form large-scale objects. One would like to understand their assembly and how this assembly is regulated. As a first step, we investigate the phase transition structure of a class of simple cooperative assembly systems. We characterize which of these systems have a (probabilistic) equilibrium and find an explicit form for their local energy (§2). We obtain, under additional limitations on cooperativity, the average dynamics of some partial observables (§4). Combining both steps, we obtain conditions for the appearance of a large cluster (§5).

Numerous intracellular molecular systems (post-synaptic densities, bacterial chemotaxis sensors, or focal adhesion complexes) form extensive structures which are somewhat intermediate between traditional signalling complexes with a few components, and larger objects more familiar from statistical physics and made of a huge number of components. One would like to understand the specific laws governing the assembly and regulation of such intermediate objects and how this regulation relates to their role in processing information.

As a first step, this paper investigates a simple model of assembly systems with only two types of agents, with limited cooperativity, and derives approximate conditions for *criticality*, that is to say for the appearance of a large cluster.

We characterize which of these idealized systems have a probabilistic equilibrium and find an explicit form for their associated (local) energy functional. Then, drawing on recent ideas from rule-based model reduction [7], we obtain under conditions limiting the type of cooperativity further, a differential system describing the average dynamics of some partial observables appearing in the energy functional. This gives us in particular their average steady state concentrations (as implicit functions of the cooperativity parameters). By combining both steps, we obtain a criticality condition for the appearance of a large cluster. The condition uses the mean reproduction rate of a suitable branching process (as is standard in the statistical physics literature [5]).

This improves on earlier similar work which considered simpler agents with no ability to cooperate [10]. Simulations for our idealized cooperative model can be set up easily using rule-based modeling languages; §3 presents some of our numerical experiments, using a recent implementation KaSim of the rule-based

* This paper is an invited contribution to the Proceedings of the 17th International Conference on DNA Computing and Molecular Programming (2011).

** Corresponding author: vdanos@inf.ed.ac.uk.

language Kappa. An example model is described in the Appendix. Simulation files and further useful information can be found here.

1 The cooperative assembly model

The assembly model we study considers only two types of agents or nodes, A and B , although some of the conclusions we attain are clearly generalizable. We suppose each agent is equipped with a set of identical sites, namely, agents of type A have v_A sites of type a , and agents of type B have v_B sites of type b . This is only to make the model tractable and carries no deep meaning.

Sites of type a and b can bind according to rules which we describe below, and we call the resulting objects *site graphs*, to insist that linking is done via the use of sites and not directly on nodes, as with usual graphs. One can think of sites as domains of proteins, or DNA domains, or other media as used in DNA-based programmable chemistry, as long as elements glue along well-defined interfaces (see eg [13]) and steric problems are not a concern. Saying that sites are identical means that the rules will not distinguish various occurrences of a or b . We refer to v_A, v_B as the *valences* of A, B , and assume $v_A, v_B > 0$. We write n_A for the total number of agents of type A , and n_B for that of type B , assume $n_A, n_B > 0$; we write $\mathcal{G}(n_A, n_B)$ for the set of bipartite site graphs with n_A, n_B nodes of type A, B , and x_0 for the unique site graph in $\mathcal{G}(n_A, n_B)$ which has no links. This site graph will serve as our origin and we will assign energy zero to it.

The rules are as follows. For each pair $0 \leq i < v_A, 0 \leq j < v_B$, we assume a reversible association rule, $r(i, j)$, whereby one can bind a pair of nodes of types A, B , and respective degrees i and j , via a pair of free sites a, b ; or unbind an already existing bond. We write $\gamma_{i,j}^+, \gamma_{i,j}^- > 0$ respectively for the association and dissociation rate of $r(i, j)$. These are the rates at which the corresponding rule is applied to one instance. This defines a continuous-time Markov chain in the usual fashion. We set $\Gamma_{i,j} = \gamma_{i,j}^- / \gamma_{i,j}^+ \in \mathbb{R}_+$, and write Γ for the associated $v_A \times v_B$ matrix.

As sites of the same type cannot bind, every site graph reachable from x_0 is bipartite (Fig. 3), and so is in $\mathcal{G}(n_A, n_B)$. Conversely, every site graph in $\mathcal{G}(n_A, n_B)$ can be reached from the totally disconnected state x_0 (hence from any other one, by symmetry of the transitions) by creating all the needed connexions. Hence, the continuous-time Markov chain underpinning the dynamics is irreducible and positive-recurrent, and, its state space $\mathcal{G}(n_A, n_B)$ being finite, it has a unique invariant probability.

Note that, as soon as $v_A, v_B \geq 2$, the dynamics can build polymers, meaning connected site graphs the size of which is only bounded by the total number of available nodes $N = n_A + n_B$ (Fig. 3).

Depending on the definition of an instance of a rule, the set of graphs reachable from x_0 might include double AB bonds, which we call ‘alkenes’ in this paper. The model, presented in the appendix, allows double bonds because the Kappa rules are shorter to write, but as this complicates things a bit on the theory side, we suppose here that rules can only bind sites from nodes that are

not yet sharing an edge. We will see in §3 that the discrepancy in simulations is negligible (Fig. 4).

2 Equilibrium

Throughout the paper we write $\log(x)$ for the *natural* logarithm, $\binom{n}{m}$ for the binomial coefficient, $[m; n] = m! \binom{n}{m}$ for the number of injections of m into n , where $m, n \geq 0$, and $\binom{n}{n_1 \dots n_k}$ with $n_i \geq 0$, $n_1 + \dots + n_k = n$ for the multinomial coefficient.

Recall Stirling's formula $\log n! \sim n \log n - n$, from which follows that:

$$\log \binom{n}{n_1 \dots n_k} \sim nH(n_i/n)$$

where $H(p) = -\sum_i p_i \log p_i$ is the Shannon entropy of a probability p with finite support. We will use this later in this section.

Our first step is to characterize the models for which the invariant probability is an equilibrium (aka has detailed balance), as when it is the case, this will give us a computational handle on the behaviour of the system.

Proposition 1 *Γ has an equilibrium iff $\Gamma = \Gamma_A \Gamma_B^t$ for some $\Gamma_A \in \mathbb{R}_+^{v(A)}$, $\Gamma_B \in \mathbb{R}_+^{v(B)}$.*

in other words if for $0 \leq i < v_A$, $0 \leq j < v_B$, one has $\Gamma_{i,j} = \Gamma_i \Gamma_j$

Proof. Consider two states x, y in $\mathcal{G}(n_A, n_B)$ such that there exists a (one-step) transition from x to y (and therefore from y to x , by reversibility of rules).

Without loss of generality, suppose y is obtained from x by adding a bond between nodes u, v of respective types A and B .

Both u, v have a definite degree in x , say i and j , and there is therefore only one rule the application of which underlies this transition, namely $r(i, j)$. The ratio of the backward and forward rates is given by:

$$\rho(x, y) = \frac{\Gamma_{i,j}}{(v_A - i)(v_B - j)} \tag{1}$$

Indeed, in order to bind nodes u, v of x , one has to choose one of the $v_A - i$ free sites of u , and the same for v . As $i < v_A$, and $j < v_B$ the above formula is always defined.

By connectedness of the transition graph, over $\mathcal{G}(n_A, n_B)$, we know that there is at most one energy function V which describes the equilibrium (it might not exist but it is unique), up to an additive constant.

Setting $V(x_0) = 0$ (which one can always do as energy is defined up to an additive constant), we know that if V exists, it must verify:

$$V(x) = \sum_{(z, z') \in \phi: x_0 \rightarrow x} \log \rho(z, z') \tag{2}$$

where ϕ is a sequence of successive transitions leading from x_0 to x . For V to be consistent, we need this definition to be independent of the choice of ϕ . It

is enough to show that two successive bindings on the same node u obtain the same ΔV , as this is the only case where two instances of rules interact in a non trivial way (up to reversibility).

So, suppose $u : A$ has degree i in x and gets bound successively to $v_1 : B$ of degree j_1 , to obtain the intermediate y , and then to $v_2 : B$ of degree j_2 , to obtain z . (As we do not allow alkene formation, $v_1 \neq v_2$.)

The product of the rate ratios $\rho(x, y_1)\rho(y_1, z)$ along the two transitions is:

$$\frac{\Gamma_{i,j_1} \Gamma_{i+1,j_2}}{(v_A - i)(v_B - j_1)(v_A - i - 1)(v_B - j_2)}$$

Path-independence forces the above to be constant under exchanging the roles of v_1 and v_2 , which leads to the following simple constraint (note that the denominator does not depend on the order):

$$\Gamma_{i,j_1} \Gamma_{i+1,j_2} = \Gamma_{i,j_2} \Gamma_{i+1,j_1}$$

and by symmetry (exchanging the roles of A and B):

$$\Gamma_{i_1,j} \Gamma_{i_2,j+1} = \Gamma_{i_1,j+1} \Gamma_{i_2,j}$$

It is easy to see that the above two equations hold iff:

$$\Gamma_{0,0} \Gamma_{i,j} = \Gamma_{i,0} \Gamma_{0,j}$$

and the statement follows, eg with $\Gamma_A(i) = \Gamma_{i,0} \Gamma_{0,0}^{-\frac{1}{2}}$, $\Gamma_B(j) = \Gamma_{0,j} \Gamma_{0,0}^{-\frac{1}{2}}$. \square

There are a few remarks worth making here: first, allowing for alkenes would not add any further constraints on the rates; second, the equilibrium condition is reminiscent of the condition derived on an earlier paper for the validity of a static analysis of reachability based on views [2] - it forces a simple multiplicative form of the dependency of Γ s in the degrees i, j ; third, the condition above does not constrain rate constants but only their ratios $\Gamma_{i,j}$, thus, it is always possible to multiply jointly $\gamma_{i,j}^+$ and $\gamma_{i,j}^-$ by an arbitrary positive scalar, without altering the equilibrium or its existence.

We suppose from now on that Γ satisfies the equilibrium condition.

The proof above, in the case the equilibrium condition is fulfilled, gives an explicit expression for the energy:

Proposition 2 *Supposing Γ has an equilibrium, the underlying energy is:*

$$V(x) = \sum_{u \in x} \log \frac{\prod_{0 \leq i < d(u)} \Gamma_{\tau(u)}(i)}{[d(u); v_{\tau(u)}]} \quad (3)$$

with $\tau(u) \in \{A, B\}$ the type of node u in x , $v_{\tau(u)}$ its valence, and $d(u)$ its degree.

Proof. We prove the formula by induction on the number of bonds in x .

If there are none, then $x = x_0$ and $V(x_0) = 0$.

Suppose now this is true for x , and add a bond in x between nodes $u : A$ and $v : B$ with respective degrees i, j . The difference of energy incurred by this addition is:

$$V(y) - V(x) = \log \frac{\Gamma_{\tau(u)}(d(u))}{(v_{\tau(u)} - d(u))} + \log \frac{\Gamma_{\tau(v)}(d(v))}{(v_{\tau(v)} - d(v))}$$

which is indeed equal to $\log \rho(x, y)$ as in Eq.1. \square

As said earlier, the dynamics described by Γ will drive the system to the invariant probability with support $\mathcal{G}(n_A, n_B)$ associated to the energy $V(x)$:

$$p_V(x) = \frac{e^{-V(x)}}{\sum_{y \in \mathcal{G}(n_A, n_B)} e^{-V(y)}}$$

Note that the above calculations hold more generally for any set of homogeneous agents as long as their association/dissociation rules are limited to depend only on the degree of the bindees.

2.1 Energy, Entropy

In this subsection, we work out some constraints on the dynamics implied by the specific format of the potential V , which we have constructed above.

We write $[A_i; x]$ for the number of *embeddings* into x of a single agent A with degree i , $A_i(x) := [A_i; x]/i!$ for the number of *occurrences* of A_i in x , and $\mathbf{A}(x), \mathbf{B}(x)$ for the associated *degree vector*; that is to say $\mathbf{A}(x) \in \mathbb{R}^{v(A)}$ has i th component $A_i(x)$, and similarly for $\mathbf{B}(x) \in \mathbb{R}^{v(B)}$.

It is easy to see that we can rewrite Eq.3 for V as:

$$V(x) = \sum_{0 < i \leq v_A} \epsilon_A(i) A_i(x) + \sum_{0 < i \leq v_B} \epsilon_B(i) B_i(x) \quad (4)$$

with $\epsilon_\tau \in \mathbb{R}^{v(\tau)}$ defined component-wise as:

$$\epsilon_\tau(i) := \sum_{0 \leq j < i} \log \frac{\Gamma_\tau(j)}{v_\tau - j}$$

We can write succinctly $V(x) = \langle \epsilon_A, \mathbf{A}(x) \rangle + \langle \epsilon_B, \mathbf{B}(x) \rangle$. As a consequence, x s with the same $\mathbf{A}(x)$ and $\mathbf{B}(x)$ are equally likely at equilibrium. In particular swapping links, or permuting nodes of the same type, leaves the energy unchanged. The set of such graphs form an equivalence class. The distinguished state x_0 belongs to a singleton class (for generic values of $\epsilon_\tau(i)$ that is) which is labelled by the degree vectors $\mathbf{0}, \mathbf{0}$.

As $\sum i A_i(x) = \sum j B_j(x)$ is the total number of bound a or b sites in x , the log of the cardinality of the class of x , or the class *entropy*, is:

$$S(x) = \log(\sum i A_i(x))! + \log \binom{n_A}{\mathbf{A}(x)} + \log \binom{n_B}{\mathbf{B}(x)} \quad (5)$$

This way to look at energy, invites one to think of $\epsilon_\tau(i)$ as setting the penalty/reward for a node of type τ to acquire degree i . We mention in passing that the idea of using a flexible syntax of *local* energy functionals in modeling biomolecular interactions has a long tradition, which was recently revived in Refs. [12, 14]. See Ref. [3] for a similar investigation in the case of ordinary chemical reactions (aka Petri nets).

2.2 Cooperativity

The first question we can ask is that of *cooperativity*. The assembly rules manifest cooperativity if adding a bond to an already busy agent results in a smaller change of potential ΔV than would adding it to an agent less busy [16]. Adding a bond on a pair A_i, B_j , $i < v_A$, $j < v_B$, results in a difference of potential:

$$\begin{aligned}\Delta V &= \epsilon_A(i+1) - \epsilon_A(i) + \epsilon_B(j+1) - \epsilon_B(j) \\ &= \log \Gamma_A(i)\Gamma_B(j)/(v_A-i)(v_B-j)\end{aligned}$$

as follows from the definition above.

Hence it is energetically more advantageous to add an (i, j) bond than an (i', j') iff:

$$\frac{\Gamma(i, j)}{\Gamma(i', j')} = \frac{\Gamma_A(i) \Gamma_B(j)}{\Gamma_A(i') \Gamma_B(j')} \leq \frac{v_A - i}{v_A - i'} \frac{v_B - j}{v_B - j'}$$

The conclusion is that it is *not* enough, in order to have cooperativity, to suppose that $\Gamma_\tau(i)$ is a decreasing function of the degree i (that would be $\Gamma_\tau(i)/\Gamma_\tau(i') \leq 1$ when $i > i'$); instead we need to assume that $\Gamma_\tau(i)$ decreases faster than $(v_\tau - i)/(v_\tau - i')$ (which is already ≤ 1 , again supposing $i > i'$). In particular, a system of rules where Γ s are constant will be anti-cooperative, if slightly. However, as $\partial v((v-i)/(v-i')) = (i-i')/(v-i')^2 \geq 0$, when $i > i'$, the larger v , the weaker the effect.

2.3 Bond displacement

Here is another question we can investigate. Suppose x is in $\mathcal{G}(n_A, n_B)$, and u_1, u_2 are nodes in x of the same type τ , and with respective degrees $i_1 > 0$, $i_2 < v_\tau$. Then, we can define a new site graph y by giving one link of u_1 (donor) to u_2 (acceptor).

Lemma 1 *Suppose y is obtained from x by bond displacement, then, the difference of potential $\Delta V := V(y) - V(x)$ and entropy $\Delta S := S(y) - S(x)$ are:*

$$\Delta V = \log \frac{\Gamma_\tau(i_2)}{\Gamma_\tau(i_1 - 1)} + \log \frac{v_\tau - i_1 + 1}{v_\tau - i_2}$$

$$\begin{aligned}\Delta S &= \log \tau_{i_1}(x)\tau_{i_2}(x) - \log (\tau_{i_1-1}(x) + 1)(\tau_{i_2+1}(x) + 1) \quad \text{if } i_2 + 1 \neq i_1 - 1 \\ &= \log \tau_{i_1}(x)\tau_{i_2}(x) - \log (\tau_k(x) + 2)(\tau_k(x) + 1) \quad \text{if } k = i_2 + 1 = i_1 - 1\end{aligned}$$

where $i_1 > 0$, $i_2 < v_\tau$ are the degrees of the donor and acceptor, and $\tau_i(x)$ is the number of agents of type τ with degree i in x .

Proof. Bond displacement preserves the total number of links, as well as the degrees of nodes not of type τ . So, in Eq.5, only the τ -side multinomial in S is changed. The expression for ΔV is an easy computation. \square

From the expression for ΔS , we see that if the target degrees $i_1 - 1$, $i_2 + 1$ have lower counts in x than the source degrees i_1 and i_2 , then $\Delta S \geq 0$, meaning the class cardinality increases. Purely based on this entropic contribution, bigger classes, which get a higher probability, will tend to have uniform degree distribution.

What about the energy contribution? Supposing $\Gamma_\tau(i) = \Gamma$ is constant, we see from the expression for ΔV above, that:

$$\begin{aligned}\Delta V &= \log(v_\tau - i_1 + 1) - \log(v_\tau - i_2) \\ \Delta V \leq 0 &\text{ iff } i_2 < i_1\end{aligned}$$

Hence states where degree differences between nodes of type τ are > 1 will be penalized. Purely based on this energy contribution (and $\Gamma_\tau(i)$ constant), graphs where the degree vector is concentrated will be favoured at equilibrium.³

The question now is which of the opposite forces wins asymptotically in $N = n_A + n_B$ the total number of nodes.⁴

2.4 A rough estimate of $\mathbf{B}(x)$ given $\mathbf{A}(x)$

Suppose, to lighten the notation, that the Γ -constant type is B . Suppose further that the $\mathbf{A}(x)$ distribution is *fixed*, which also fixes the total number of links as $\sum_i i B_i(x) = \sum_j j A_j(x)$. We will discuss this latter assumption, which is key, later in §3.

Write $p_i(x) := B_i(x)/n_B$ for the probability that a randomly chosen B in x has degree i , and $q_i(x) := \sum_{i \leq k \leq v_B} p_k(x)$ for the probability that such a B has degree $\geq i$ (a non-increasing sequence).

The asymptotic entropic contribution, given $\mathbf{A}(x)$, is, up to an additive constant:

$$S(\mathbf{B}(x)) = \binom{n_B}{\mathbf{B}(x)} \sim n_B H(p) \leq n_B \log v_B$$

On the other hand the energy contribution, again given $\mathbf{A}(x)$, and using the Γ_B constant assumption, is, up to an additive constant:

$$\begin{aligned}-V(\mathbf{B}(x)) &= \sum_{0 \leq i \leq v_B} B_i(x) \log[i; v_B] \\ &= n_B \sum_{0 \leq i \leq v_B} p_i(x) \log[i; v_B] \\ &= n_B \sum_{1 \leq k \leq v_B} q_k(x) \log(v_B - k + 1) \\ &= q_1(x) n_B \log v_B + q_2(x) n_B \log(v_B - 1) + \dots\end{aligned}$$

³ As $\partial v_\tau \exp(-\Delta V) = (i_1 - 1 - i_2)/(v - i_1 + 1)^2$ has the same sign as ΔV , higher valences will increase the degree distribution concentration of nodes of type τ .

⁴ As an aside, we can observe that the ΔV of bond displacement proves that the distribution of degrees on τ s with constant Γ_τ is not a per site independent distribution - given that all other sites of a node u are free, it will be far likelier that a given site of u is bound.

The likeliest $\mathbf{B}(x)$ will minimize the *free energy* $V(\mathbf{B}(x)) - S(\mathbf{B}(x))$, given the total number of links. One sees that, asymptotically, and assuming $q_1(x) \sim 1$, which is definitely going to be the case near criticality, the entropy contribution is trumped by the energy one. See Figs. 1, 2 for concrete examples.

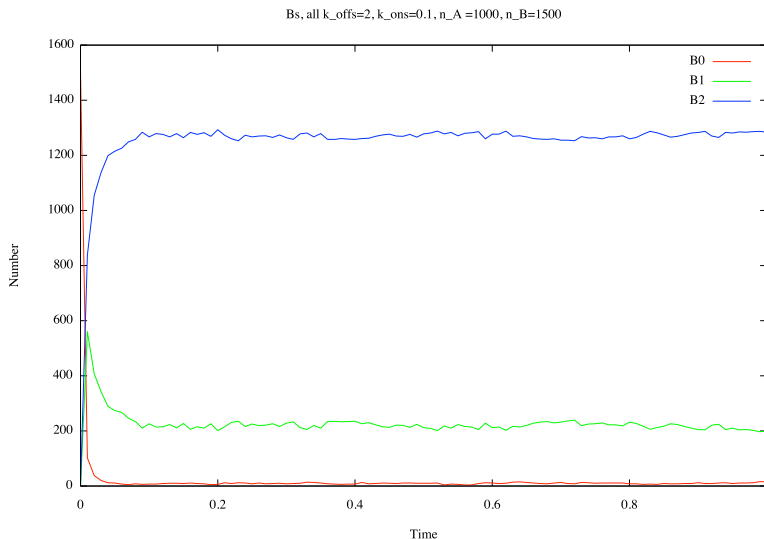


Fig. 1. We observe $\mathbf{B}(x)$ along a simulation with parameters: $\gamma_{i,j}^- = 2$, $\gamma_{i,j}^+ = 0.1$, $n_A = 1000$, $n_B = 1500$, $v_A = 3$, $v_B = 2$. Note that B_0 rapidly becomes negligible.

So, in this constant case, it seems reasonable (and all the more reasonable when v_B is high) to *neglect* the entropic contribution on the B -side, given $\mathbf{A}(x)$, and to approximate the equilibrium distribution for the degree of nodes of type B by the following *two-valued distribution* which minimizes the potential V given a fixed number of links $\sum iB_i$, which is described in the next lemma.

Lemma 2 *The 2-valued degree vector \mathbf{B} defined as:*

$$\begin{aligned} \mathbf{B}_q &= n_B - r \\ \mathbf{B}_{q+1} &= r \end{aligned}$$

with $\lambda = qn_B + r$, $0 \leq r < n_B$ minimizes V , given a number of links λ .

Proof. First, any > 3 -valued degree vector \mathbf{B} has a non-locally minimal V , as one can pick a (link) donor in the top band and an acceptor in the bottom one, and $\Delta V \leq 0$ from the preceding lemma. So a local minimum is 2-valued at most. Any such degree vector must verify the following two independent equations:

$$\begin{aligned} n_B &= n_q + n_{q+1} \\ \lambda &= qn_q + (q+1)n_{q+1} \end{aligned}$$

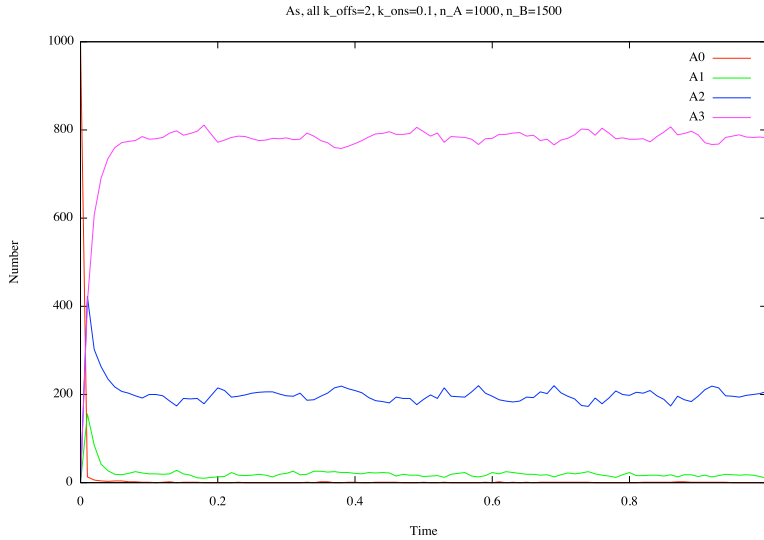


Fig. 2. We observe now $\mathbf{A}(x)$ along a simulation with the same parameters as above in Fig. 1. Again A_0 rapidly becomes negligible, and we see that the distribution concentrates on degrees 2 and 3, as in the estimate of §2.4.

The determinant is 1 so the equations are indeed independent. So there is a unique solution (which will therefore be a global minimum). For the \mathbf{B} above, the first relation clearly holds and for the second, we have $q(n_B - r) + (q+1)r = qn_B + r = \lambda$, so it holds as well. \square

This V -minimal assignment obtained by Euclidean division of the total number of bonds by the total number of nodes, will serve as an estimate of the true one.

Encouragingly, simulations in Figs. 1, 2, where both Γ_A, Γ_B are constant seem to agree with our estimate. On the A -side, we see that A_s concentrates on degrees 2 and 3, and the sampled value $\mathbf{A} = (0, 10, 210, 780)$ compares well with the estimate $\hat{\mathbf{A}} = (0, 0, 230, 770)$, which we can use since Γ_A is constant. On the B -side, B_s concentrate on degrees 1 and 2, and we get $\mathbf{B} = (15, 201, 1284)$ which again compares well with $\hat{\mathbf{B}} = (0, 231, 1269)$, if slightly less so, which is expected as the valence of B is lower. This estimate is not at all rigorous, but caution notwithstanding, we will use it in §5.

3 Simulations

We now look more closely at simulations of some of our assembly systems. As in §2, we choose $v_A = 3, v_B = 2$ (enough to assemble polymers of unbounded size given unboundedly many nodes). Perfect matchings are reachable only if

the stoichiometry is $2n_A = 3n_B$; this is the stickiest the system can get, all other things being equal, as in this case all sites are used to form a bond. Eg if we pick $n_A, n_B = 100, 150$, we get an equilibrium probability on $\mathcal{G}(100, 150)$ of which a typical sample looks as in Fig. 3, indeed almost a perfect matching.

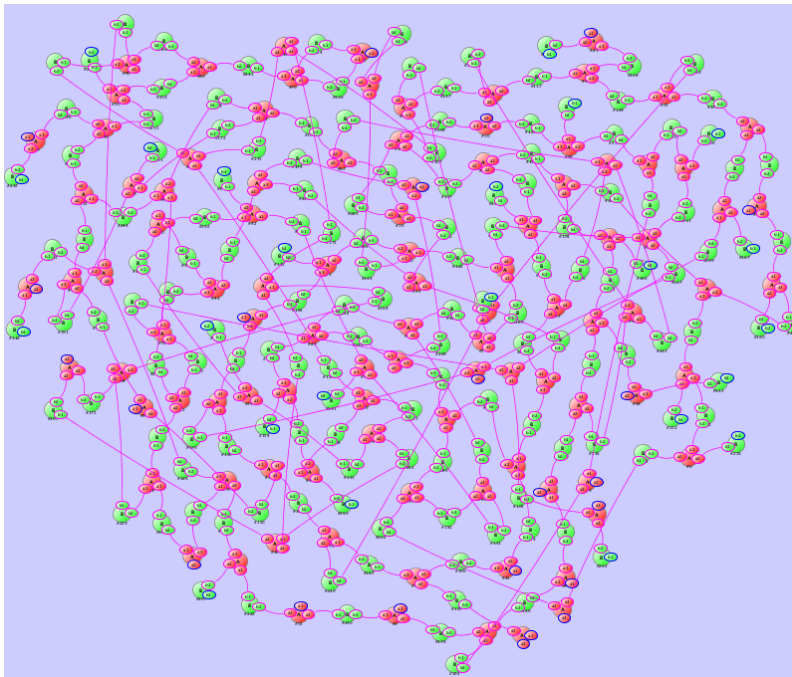


Fig. 3. A long-term snapshot from the above system with parameters: $\gamma_{i,j}^- = 2, \gamma_{i,j}^+ = 1$, $n_A = 100, n_B = 150$. One sees that the graph alternates. With this set of parameters, we are well past criticality, as we have just one connected site graph.

We pick parametrizations where γ^+ is constant and γ^- only depends on A 's degree. Hence, $\Gamma(i, j) = \Gamma(i)$ does not depend on j and the conditions for equilibrium given in §2 are clearly satisfied. As Γ_B is constant, the additional assumption introduced in §2.4 also is. We choose $\Gamma_A(i)$ as a monotonic function of i to simulate both cooperativity and anti-cooperativity (with the provision discussed in §2.2). We see in the resulting simulations in Fig. 5 that alkenes are extremely rare, and that, unsurprisingly, the cooperative case is 'stickier' (has a higher edge count).

Another thing one sees, is that, clearly, there is no good way to talk about a deterministic steady state of such systems, as for large populations, the state space accessible to probabilistic equilibrium is always increasing. This does not prevent, however, *local* features of this distribution to take quasi-deterministic values in the limit. Indeed, looking again at Fig. 5, one notices that, numerically,

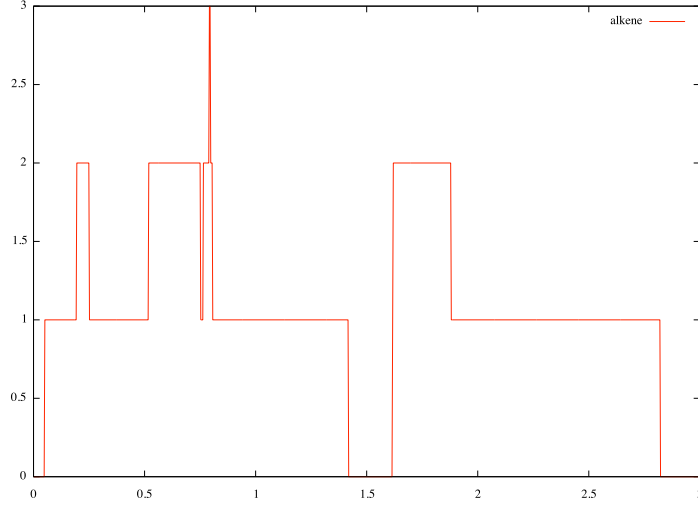
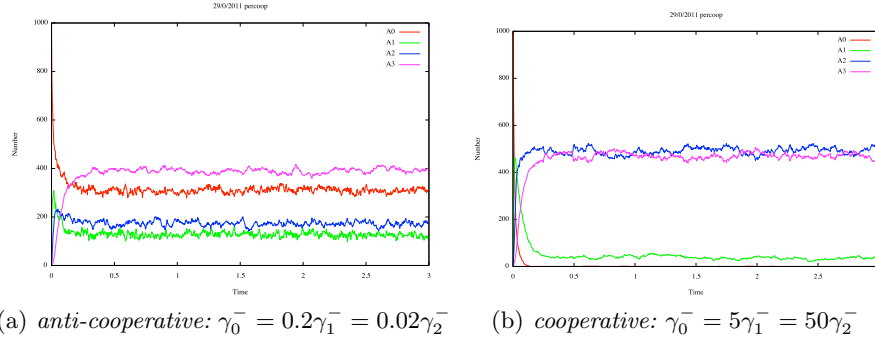


Fig. 4. The total alkenes stays very low with cooperative parameters: $\gamma_0^- = 5\gamma_1^- = 50\gamma_2^-$, $n_A = 1000$, $n_B = 1500$, $\gamma_{i,j}^+ = 0.1$.



(a) anti-cooperative: $\gamma_0^- = 0.2\gamma_1^- = 0.02\gamma_2^-$

(b) cooperative: $\gamma_0^- = 5\gamma_1^- = 50\gamma_2^-$

Fig. 5. Parameters specific to the cooperative and anti-cooperative cases are indicated in the caption. In both cases, $\gamma^+(i) = 0.1$ is kept constant, and one only varies $\gamma^-(i)$, while $n_A = 1000$, $n_B = 1500$.

$\mathbf{A}(x)$ reaches a quasi-deterministic value. As the number of nodes increases, the effect is reinforced (not shown).

So it is natural to assume constant values for large times for $\mathbf{A}(x)$ and $\mathbf{B}(x)$. The question remains as to how to get these values. One obvious way to do this is by simulation, measuring their average long-term values. But, this limits us to concrete examples. In the next section, assuming Γ_B constant, we show how to derive an equation, the solution of which gives the steady state value of $\mathbf{A}(x)$.

4 Deterministic equilibrium for $\mathbf{A}(x)$

Our next step is to write down a differential equation which describes the average evolution of the $\mathbf{A}(x)$ and $\mathbf{B}(x)$. This is part of a generic process of *fragmentation* of rule-based models described in Ref. [1, 6, 7]. The presentation in §4.1-2 is self-contained. In this simple case, there is a shortcut to the answer, so the reader not interested in the machinery can jump directly to §4.3.

4.1 Fragmentation

We have to consider consumption terms caused by the forward rules $r(i, j)$ and the backward ones $r^*(i, j)$, as well as their production terms. We will write simply A_i, B_j (where previously we wrote $A_i(x), B_j(x)$), and $A_i \circ B_j$ for the ‘bond observable’, namely a pair of an A_i and a B_j joined by a (unique) bond. This implies $i, j > 0$, as in our convention, indexes i, j include the bond itself (eg, $A_1 \circ B_1$ denotes a bond between an A and a B which have no other bond).

We can decompose the event horizon of $x \in \mathcal{G}(n_A, n_B)$ (that is to say the set of all events that are possible at x) depending on:

- the rule the event is a event of, and
- whether the event increases/decreases the number of A_i s (indicated by a \pm sign below).

The second column in the tables below records the count of the total number of consumption/productions pairs where an A_i occurrence and an event intersect, meaning that the event modifies the said occurrence. Note that by convention here we produce/consume occurrences (events are defined as in §2’s definition of applying a rule) not matchings, as this is simpler in our symmetric setting.

For the consumptions:

Rule	# collision pairs	conditions on i, j
$r(i, j)$	$(v_A - i)(v_B - j)A_i B_j$	$0 \leq i, j < v_A, v_B$
$r^*(i - 1, j - 1)$	$A_i \circ B_j$	$0 < i, j \leq v_A, v_B$

For the productions:

Rule	# production pairs	conditions on i, j
$r(i - 1, j - 1)$	$(v_A - i + 1)(v_B - j + 1)A_{i-1} B_{j-1}$	$0 < i, j \leq v_A, v_B$
$r^*(i, j)$	$A_{i+1} \circ B_{j+1}$	$0 \leq i, j < v_A, v_B$

As the number of collision/production pairs involving A_i multiplied by the rule rate constant gives the average number of modifications of A_i per time unit, we can write the general differential equation for A_i :

$$\begin{aligned}
A'_i = & -\mathbf{1}_{i < v_A} \sum_{0 \leq j < v_B} \gamma_{i,j}^+ (v_A - i)(v_B - j) A_i B_j \\
& - \mathbf{1}_{i > 0} \sum_{0 < j \leq v_B} \gamma_{i-1,j-1}^- A_i \circ B_j \\
& + \mathbf{1}_{i > 0} \sum_{0 < j \leq v_B} \gamma_{i-1,j-1}^+ (v_A - i + 1)(v_B - j + 1) A_{i-1} B_{j-1} \\
& + \mathbf{1}_{i < v_A} \sum_{0 \leq j < v_B} \gamma_{i,j}^- A_{i+1} \circ B_{j+1}
\end{aligned} \tag{6}$$

We get similar equations for the B_j s, but we are not interested in those.⁵

4.2 Additional assumption

The problem with the equations above is that they do not form a *self-consistent* system. Some terms require the knowledge of the ‘bond observables’ $A_i \circ B_j$ which one can *a priori* only infer from the knowledge of x . Following the general fragmentation procedure at this stage, would need writing a similar ODE for the $A_i \circ B_j$ s. This is possible. But the problem is that one can now have an event centered on a given bond between u, v of type A, B (whether to create it or to delete it) which intersects other bond observables attached to u and v . Thus, the time derivative of $A_i \circ B_j$ will contain ternary terms of the form $A_i \circ B_j \circ A_k$, etc. It is easy to show that this procedure generates unboundedly many observables. So if we want some handle on the eventual mean values of \mathbf{A}, \mathbf{B} , we need to do something else.

Suppose $\gamma_{i,j}^- = \gamma_i^-$, and $\gamma_{i,j}^+ = \gamma_i^+$ do not depend on j , we can rewrite the ODE above self-consistently:

$$\begin{aligned}
A'_i = & \mathbf{1}_{\{i < v_A\}} \cdot (\gamma_i^- (i + 1) A_{i+1} - \gamma_i^+ (v_A - i) A_i n_b^f) \\
& + \mathbf{1}_{\{i > 0\}} \cdot (-\gamma_{i-1}^- i A_i + \gamma_{i-1}^+ (v_A - i + 1) A_{i-1} n_b^f)
\end{aligned} \tag{7}$$

where we have written n_b^f for the the *number of free sites of type b* in x (not to be confused with the number of free B 's, written B_0) which satisfies:

$$n_b^f = v_B n_B - \sum_{0 \leq j \leq v_B} j B_j = v_A n_A - \sum_{0 \leq i \leq v_A} i A_i \tag{8}$$

This simplified differential system, indexed by $0 \leq i \leq v_A$ is now indeed self-consistent as every variable, including n_b^f which using Eq.8 can also be expressed

⁵ For the reader familiar with fragmentation, as the graph is bipartite there is no approximation in expressing $[A_i, B_j; x]$ as $[A_i; x] \times [B_j; x]$.

in terms of the A_i s, and thus, there is no longer a need to look back at x to understand the implied dynamics (even the B_j s are not needed).

To see in details how the terms of the equation above simplify, consider the following:

$$\begin{aligned}
\sum_{0 \leq j < v_B} \gamma_{i,j}^+(v_A - i)(v_B - j)A_i B_j &= \gamma_i^+(v_A - i)A_i n_b^f \\
\sum_{0 < j \leq v_B} \gamma_{i-1,j-1}^- A_i \circ B_j &= \gamma_{i-1}^- \sum_{0 < j \leq v_B} A_i \circ B_j = \gamma_{i-1}^- i A_i \\
\sum_{0 < j \leq v_B} \gamma_{i-1,j-1}^+(v_A - i + 1)(v_B - j + 1)A_{i-1} B_{j-1} &= \gamma_{i-1}^+(v_A - i + 1)A_{i-1} n_b^f \\
\sum_{0 \leq j < v_B} \gamma_{i,j}^- A_{i+1} \circ B_{j+1} &= \gamma_i^- \sum_{0 \leq j < v_B} A_{i+1} \circ B_{j+1} = \gamma_i^- (i + 1)A_{i+1}
\end{aligned}$$

Note that our new assumption, which will hold onwards, $\gamma_{i,j}^\pm = \gamma_i^\pm$, evidently entails the equilibrium existence condition of Th.1. It would be enough to suppose $\gamma_{i,j}^- = \gamma_i^-$ to obtain self-consistency, but with a more complex ODE, which does not look very tractable. It is worth noting that our additional assumption breaks the symmetry between A s and B s; in particular, we are not able to write *at the same time* a self-consistent ODE system for the B_j s, unless we suppose that $\gamma_{i,j}^-$ does not depend on i either (which takes us back to the simpler case treated in Ref. [4]).

4.3 Deterministic steady state

Putting everything together we obtain:

Proposition 3 *The deterministic steady state value of A_i , $0 \leq i \leq v_A$, assuming $\gamma_{i,j}^\pm = \gamma_i^\pm$, is given by the following system of equations with parameters n_A , n_B , and Γ_i , $0 \leq i \leq v_A$:*

$$\begin{aligned}
n_A &= \sum_{0 \leq i \leq v_A} A_i \\
n_b^f &= v_B n_B - \sum_{0 \leq i \leq v_A} i A_i \\
A_i &= \binom{v_A}{i} \prod_{0 \leq k < i} \Gamma_k^{-1} \cdot (n_b^f)^i \cdot A_0
\end{aligned} \tag{9}$$

Proof. If we write $\alpha_i^- = \gamma_i^-(i + 1)$, and $\alpha_i^+ = \gamma_i^+(v_A - i)n_b^f$, and set $A_i' = 0$ in Eq.7, we get the following systems of equations:

$$\begin{aligned}
\alpha_0^- A_1 &= \alpha_0^+ A_0 \\
\alpha_{i-1}^- A_i + \alpha_i^+ A_i &= \alpha_i^- A_{i+1} + \alpha_{i-1}^+ A_{i-1} \\
\alpha_{v_A-1}^- A_{v_A} &= \alpha_{v_A-1}^+ A_{v_A-1}
\end{aligned}$$

which implies $A_{i+1} = \alpha_i^+ / \alpha_i^- A_i = (v_A - i) / (i + 1) \cdot \Gamma_i^{-1} \cdot n_b^f \cdot A_i$ for $0 \leq i < v_A$. \square

Note that there is a *shortcut* to derive the above equations, by saying directly that, at equilibrium, one must have:

$$\gamma_i^-(i+1)A_{i+1} = \gamma_i^+(v_A - i)n_b^f A_i$$

that is to say the likelihood that an A_{i+1} -link breaks has to equal the likelihood that one is created.

We can rescale these equations by making all quantities relative to the total node population N ; this needs modifying the ‘volume’ of the system and replacing Γ_i with Γ_i/N .

Writing $a = n_A/N$, $b = n_B/N$, $a_i = A_i/N$ and:

$$K_i := \binom{v_A}{i}^{-1} \cdot \prod_{0 \leq k < i} \Gamma_k/N$$

we obtain a scale-less version of the steady state equations (thereby establishing that two models with same Γ_i/N s, a , and b will have the same a_i s, hence the same degree distributions for A s), with unknowns a_i , for $0 < i \leq v_A$:

$$K_i \cdot a_i = \left(a - \sum_{0 < k \leq v_A} a_k \right) \cdot \left(v_B b - \sum_{0 \leq k \leq v_A} k a_k \right)^i \quad (10)$$

The solutions a_i give us the mean long-term values $A_i = N a_i$, which is what we wanted. We cannot solve this system of equations in closed form, but we can solve it much faster numerically than by averaging many simulations.

5 Criticality

We would like now to understand under which conditions a giant cluster will appear at equilibrium with high probability (that is to say with a probability that tends to 1 as $N \rightarrow \infty$).

As said earlier, graphs in $\mathcal{G}(n_A, n_B)$ with the same a_i statistics are equally likely. As we assume that these a_i s are the solutions to Eq.10 above (deterministic approximation), we can interpret a_i as the probability that a given node of type A has i neighbours, and we are in the familiar situation of a fixed degree distribution random graph model (aka Molloy-Reed [5, 11]); except for the slight simplification that our graphs are all bipartite, so there is no need to condition out self-loops (we still have multiple edges or ‘alkenes’ which we neglect).

Let us write μ for $\sum i a_i = \sum j b_j$. We can think of the exploration of a connected component in terms of a branching process. We start from a bound b , which is linked to an A_i with probability α_i , and therefore gives us $(i-1)$ more bound a s to follow (because we just used one); each of these a s give a B_j with probability β_j , and so $(j-1)$ new bound b s.

Now α_i is proportional to $i A_i$, because an A node is picked with a probability proportional to its degree. In other words, $\alpha_i := i A_i / \sum_i i A_i = i a_i / \mu$, and, likewise, $\beta_j := j B_j / \sum_j j B_j = j b_j / \mu$. So the mean number of bound b sites discovered, starting with one, and completing one cycle is given by:

$$\nu = \sum_{i \leq v_A, j \leq v_B} a_i b_j i(i-1)j(j-1) / \mu^2 = \sum_{i \leq v_A} a_i i(i-1) / \mu \cdot \sum_{j \leq v_B} b_j j(j-1) / \mu$$

Introducing $p(A_i) = Na_i/n_A$, $p(B_j) = Nb_j/n_B$, the probabilities that an A has degree i , and a B degree j , one can rewrite the above as:⁶

$$\nu = \frac{N^2}{\mu^2 n_A n_B} \sum_{i \leq v_A} p(A_i) i(i-1) \cdot \sum_{j \leq v_B} p(B_j) j(j-1) = \frac{N^2}{n_A n_B} \frac{\theta_A \theta_B}{\mu^2}$$

where θ_A, θ_B are the so-called second *factorial moments* of the A and B degree distributions.

Whenever a branching law has finite mean, we know by a classical result [8, Th 6.1 8.1] that: 1) there is a non-zero probability for the associated branching process to never extinguish iff $\nu > 1$, and 2) the population in the k th generation, here the number of bound b sites, grows geometrically as ν^k . This means that, asymptotically, the exploration described above will discover an infinite cluster if $\nu > 1$. So the criticality condition reads (using notations $a = n_A/N$, $b = n_B/N$ introduced earlier in §4.3):

$$\theta_A \theta_B > \mu^2 ab \tag{11}$$

Observe that all quantities involved are scale-less and have an interesting probabilistic interpretation. First, a and b are the probabilities that a node has respective types A and B , so $a + b = 1$ and ab is the variance of the ‘composition’ of the initial state. The higher it is, the less critical the system. Second, μ is the edge density, or half the mean degree. Finally, θ_A, θ_B are variance-like quantities that measure the noise on their the degree distributions of A and B .

The above can be rewritten:

$$(\theta_A/d_A)(\theta_B/d_B) > a^2 b^2$$

where $d_A = \sum_i i p(A_i)$ is the average degree of a node of type A . Hence, it can never be fulfilled if $v_A, v_B \leq 2$, as in this case $\theta_A/d_A, \theta_B/d_B$ are both ≤ 1 .

Be that as it may, using Eq.10, ie numerically solving it, we can compute θ_A from the a_i s. This also gives the total number of links, hence $\sum_i i B_i$. This is all we need to use the estimate of §2.4, and compute θ_B . Indeed, the estimate says that B has degree q with probability $1 - \rho$, and degree $q + 1$ with probability ρ , with $\sum_i i B_i = (q + \rho)n_B$. In which case $q(q-1) \leq \theta_B = q(q-1+2\rho) \leq q(q+1)$, with q the integer part (rounded below) of the average degree $\sum_i i B_i/n_B = \mu/b$ of a node of type B .

References

1. Danos, V., Feret, J., Fontana, W., Harmer, R., Krivine, J.: Abstracting the differential semantics of rule-based models: Exact and automated model reduction. In: Proceedings of the 25th Annual IEEE Symposium on Logic in Computer Science, LICS 2010, 11-14 July 2010, Edinburgh, United Kingdom. pp. 362–381 (2010)

⁶ Earlier in §2 we have written simply p_i for $p(B_i)$.

2. Danos, V., Feret, J., Fontana, W., Krivine, J.: Abstract interpretation of cellular signalling networks. In: et al., F.L. (ed.) VMCAI'08. LNCS, vol. 4905, pp. pp. 83–97. Springer (Jan 2008)
3. Danos, V., Oury, N.: Equilibrium and termination II: the case of Petri nets (2010), to appear in MSCS
4. Danos, V., Schumacher, L.J.: How liquid is biological signalling? *Theoretical Computer Science* 410(11), 1003–1012 (2009)
5. Durrett, R.: *Random graph dynamics*, vol. 20. CUP (2007)
6. Feret, J., Danos, V., Harmer, R., Krivine, J., Fontana, W.: Internal coarse-graining of molecular systems. *PNAS* 106(16), 6453–8 (Apr 2009)
7. Harmer, R., Danos, V., Feret, J., Krivine, J., Fontana, W.: Intrinsic Information carriers in combinatorial dynamical systems. *Chaos* 20(3), 037108 (2010)
8. Harris, T.: *The theory of branching processes*. Dover (1989)
9. Koepl, H., Hafner, M., Danos, V.: Rule-based modeling for protein-protein interaction networks - the Cyanobacterial circadian clock as a case study. In: WCSB'09 proceedings. pp. 87–91 (2009)
10. Koepl, H., Schumacher, L.J., Danos, V.: A statistical analysis of receptor clustering using random graphs. In: WCSB'09 proceedings. pp. 95–99 (2009)
11. Molloy, M., Reed, B.: A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms* 6(2-3), 161–180 (1995)
12. Ollivier, J., Shahrezaei, V., Swain, P.: Scalable rule-based modelling of allosteric proteins and biochemical networks. *PLoS Computational Biology* 6(11) (2010)
13. Qian, L., Winfree, E.: Scaling up digital circuit computation with DNA strand displacement cascades. *Science* 332(6034), 1196 (2011)
14. Saiz, L., Vilar, J.: Stochastic dynamics of macromolecular-assembly networks. *Molecular Systems Biology* 2(1) (2006)
15. Sneddon, M., Faeder, J., Emonet, T.: Efficient modeling, simulation and coarse-graining of biological complexity with nfsim. *Nature Methods* (2010)
16. Williamson, J.R.: Cooperativity in macromolecular assembly. *Nat Chem Biol* 4(8), 458–465 (Aug 2008)

A Distinguishing sites - compilation to Kappa

The model described below uses the KaSim simulator. (For a general and concise example of rule-based modelling, a possible entry point is Ref. [9].) There are other rule-based simulators such as the recent NFsim which one could use to do this just as well [15].

As KaSim does not know indistinguishable sites, in order to experiment numerically with our models, we need to understand how to design equivalent ones where all sites are locally distinguished. This is an interesting question in its own right, which we deal with now.

For the purpose of the discussion we will call site graphs where sites of a given node have distinct types *concrete*, and those where, as we have done so far, we allow many sites of a same node to be of the same type, we will call *abstract*. Two concrete graphs which only differ by permutations of sites local to nodes will be said to be *equivalent*. This defines an equivalence relation on concrete site graphs, the class of which are the abstract ones.

For definiteness, in our concrete graphs, we will choose (a_1, \dots, a_{v_A}) for the sites of nodes of type A , and (b_1, \dots, b_{v_B}) for those of nodes of type B .

In order to define the dynamics on concrete site graphs, we replace the abstract reversible rule $r(i, j)$ with a *set* $R(i, j)$ of concrete reversible rules which enumerate all possibilities in the following way. We choose first a pair of a free a and a free b to be bound by the rule, and then, among the remaining sites, we choose the i bound as , the j bound bs .

Thus $|R(i, j)| = v_A v_B \cdot \binom{v_A-1}{i} \binom{v_B-1}{j}$, where the -1 in the ‘choose’ terms comes from the fact that we already have chosen one a or b , namely the one we want to bind in the concrete rule of interest.

Note that $\sum_{i, j < v_A, v_B} |R(i, j)| = v_A v_B 2^{v_A-1} 2^{v_B-1}$, so, unsurprisingly, the concrete rule set is exponentially larger than the original one, $v_A v_B$, in the valence of A and B .

Let a pair of unconnected nodes with degrees $i < v_A, j < v_B$ be given in a concrete site graph. Exactly $(v_A - i)(v_B - j)$ forward rules in $R(i, j)$ apply to this pair, and they apply uniquely (ie with a unique match). Indeed the only choice remaining in a concrete site graph, is that of the free sites. All choices lead to different concrete graphs, but the same abstract one. Similarly, given a pair of connected nodes with degrees $i \leq v_A, j \leq v_B$, exactly one backward rule in $R(i, j)$ applies, namely the one that breaks the (unique) connexion between the nodes.

Pick x, y abstract site graphs connected via an $r(i, j)$ transition, and a concrete c above x . By the considerations above, the rate at which c jumps in y (meaning to some concrete graph in the equivalence class y) is $\gamma^+(i, j)(v_A - i)(v_B - j)$, while the reverse rate at which some c' over y jumps in x is $\gamma^-(i, j)$.

Hence, the dynamics of the concrete transition system is identical to the one we have used in our theoretical investigation (one says sometimes that the quotient is a strong stochastic bisimulation of continuous-time Markov chains), and we can safely use it to explore the behaviour of our models in a numerical way. We can now turn to the description of our concrete model.

B The numerical model

Our working model can be obtained here, the simulator KaSim here. The specific instance below is parametrized as is the cooperative model run in Fig. 5(b).

B.1 Agents, Parameters and Initial State

We start by declaring agents and the main parameters, including a fictitious volume parameter to be able to rescale simulations easily.

```
%agent: B(b1,b2)
%agent: A(a1,a2,a3)

%var: 'vol' 100
%var: 'k_on' 0.1/'vol'
%var: 'k_off' 2
%var: 'k_off_vee' 1/5 * 'k_off'
%var: 'k_off_tee' 1/50 * 'k_off'
%var: 'n_A' (1000 * 'vol')
%var: 'n_B' (1500 * 'vol')
```

Then we define the initial state (written x_0 earlier).

```
%init: 'n_A' (A(a1,a2,a3))
%init: 'n_B' (B(b1,b2))
```

B.2 Rules

Now we need the interaction rules. From the preceding subsection, and since we have assumed that rules have no dependency in the degree of B , we expect to have to write $v_A v_B 2^{v_A-1} = 24$ reversible rules (so 48 KaSim rules). Note that all rates uses the parameters defined earlier for more flexibility.

```
'b1-a1-11' B(b1), A(a1,a2!_,a3!_) -> B(b1!0), A(a1!0,a2!_,a3!_)@ 'k_on'
'b1-a1-10' B(b1), A(a1,a2!_,a3 ) -> B(b1!0), A(a1!0,a2!_,a3 )@ 'k_on'
'b1-a1-01' B(b1), A(a1,a2 ,a3!_) -> B(b1!0), A(a1!0,a2 ,a3!_)@ 'k_on'
'b1-a1-00' B(b1), A(a1,a2 ,a3 ) -> B(b1!0), A(a1!0,a2 ,a3 )@ 'k_on'

'b1 a1-11' B(b1!0), A(a1!0,a2!_,a3!_) -> B(b1), A(a1,a2!_,a3!_)@ 'k_off_tee'
'b1 a1-10' B(b1!0), A(a1!0,a2!_,a3 ) -> B(b1), A(a1,a2!_,a3 )@ 'k_off_vee'
'b1 a1-01' B(b1!0), A(a1!0,a2 ,a3!_) -> B(b1), A(a1,a2 ,a3!_)@ 'k_off_vee'
'b1 a1-00' B(b1!0), A(a1!0,a2 ,a3 ) -> B(b1), A(a1,a2 ,a3 )@ 'k_off'

'b1-a2-11' B(b1), A(a2,a1!_,a3!_) -> B(b1!0), A(a2!0,a1!_,a3!_)@ 'k_on'
'b1-a2-10' B(b1), A(a2,a1!_,a3 ) -> B(b1!0), A(a2!0,a1!_,a3 )@ 'k_on'
'b1-a2-01' B(b1), A(a2,a1 ,a3!_) -> B(b1!0), A(a2!0,a1 ,a3!_)@ 'k_on'
'b1-a2-00' B(b1), A(a2,a1 ,a3 ) -> B(b1!0), A(a2!0,a1 ,a3 )@ 'k_on'

'b1 a2-11' B(b1!0), A(a2!0,a1!_,a3!_) -> B(b1), A(a2,a1!_,a3!_)@ 'k_off_tee'
```

```

'b1 a2-10' B(b1!0), A(a2!0,a1!_,a3_ ) -> B(b1), A(a2,a1!_,a3_ )@ 'k_off_vee'
'b1 a2-01' B(b1!0), A(a2!0,a1_ ,a3!_) -> B(b1), A(a2,a1_ ,a3!_)@ 'k_off_vee'
'b1 a2-00' B(b1!0), A(a2!0,a1_ ,a3_ ) -> B(b1), A(a2,a1_ ,a3_ )@ 'k_off'

'b1-a3-11' B(b1), A(a3,a1!_,a2!_) -> B(b1!0), A(a3!0,a1!_,a2!_)@ 'k_on'
'b1-a3-10' B(b1), A(a3,a1!_,a2_ ) -> B(b1!0), A(a3!0,a1!_,a2_ )@ 'k_on'
'b1-a3-01' B(b1), A(a3,a1_ ,a2!_) -> B(b1!0), A(a3!0,a1_ ,a2!_)@ 'k_on'
'b1-a3-00' B(b1), A(a3,a1_ ,a2_ ) -> B(b1!0), A(a3!0,a1_ ,a2_ )@ 'k_on'

'b1 a3-11' B(b1!0), A(a3!0,a1!_,a2!_) -> B(b1), A(a3,a1!_,a2!_)@ 'k_off_tee'
'b1 a3-10' B(b1!0), A(a3!0,a1!_,a2_ ) -> B(b1), A(a3,a1!_,a2_ )@ 'k_off_vee'
'b1 a3-01' B(b1!0), A(a3!0,a1_ ,a2!_) -> B(b1), A(a3,a1_ ,a2!_)@ 'k_off_vee'
'b1 a3-00' B(b1!0), A(a3!0,a1_ ,a2_ ) -> B(b1), A(a3,a1_ ,a2_ )@ 'k_off'

'b2-a1-11' B(b2), A(a1,a2!_,a3!_) -> B(b2!0), A(a1!0,a2!_,a3!_)@ 'k_on'
'b2-a1-10' B(b2), A(a1,a2!_,a3_ ) -> B(b2!0), A(a1!0,a2!_,a3_ )@ 'k_on'
'b2-a1-01' B(b2), A(a1,a2_ ,a3!_) -> B(b2!0), A(a1!0,a2_ ,a3!_)@ 'k_on'
'b2-a1-00' B(b2), A(a1,a2_ ,a3_ ) -> B(b2!0), A(a1!0,a2_ ,a3_ )@ 'k_on'

'b2 a1-11' B(b2!0), A(a1!0,a2!_,a3!_) -> B(b2), A(a1,a2!_,a3!_)@ 'k_off_tee'
'b2 a1-10' B(b2!0), A(a1!0,a2!_,a3_ ) -> B(b2), A(a1,a2!_,a3_ )@ 'k_off_vee'
'b2 a1-01' B(b2!0), A(a1!0,a2_ ,a3!_) -> B(b2), A(a1,a2_ ,a3!_)@ 'k_off_vee'
'b2 a1-00' B(b2!0), A(a1!0,a2_ ,a3_ ) -> B(b2), A(a1,a2_ ,a3_ )@ 'k_off'

'b2-a2-11' B(b2), A(a2,a1!_,a3!_) -> B(b2!0), A(a2!0,a1!_,a3!_)@ 'k_on'
'b2-a2-10' B(b2), A(a2,a1!_,a3_ ) -> B(b2!0), A(a2!0,a1!_,a3_ )@ 'k_on'
'b2-a2-01' B(b2), A(a2,a1_ ,a3!_) -> B(b2!0), A(a2!0,a1_ ,a3!_)@ 'k_on'
'b2-a2-00' B(b2), A(a2,a1_ ,a3_ ) -> B(b2!0), A(a2!0,a1_ ,a3_ )@ 'k_on'

'b2 a2-11' B(b2!0), A(a2!0,a1!_,a3!_) -> B(b2), A(a2,a1!_,a3!_)@ 'k_off_tee'
'b2 a2-10' B(b2!0), A(a2!0,a1!_,a3_ ) -> B(b2), A(a2,a1!_,a3_ )@ 'k_off_vee'
'b2 a2-01' B(b2!0), A(a2!0,a1_ ,a3!_) -> B(b2), A(a2,a1_ ,a3!_)@ 'k_off_vee'
'b2 a2-00' B(b2!0), A(a2!0,a1_ ,a3_ ) -> B(b2), A(a2,a1_ ,a3_ )@ 'k_off'

'b2-a3-11' B(b2), A(a3,a1!_,a2!_) -> B(b2!0), A(a3!0,a1!_,a2!_)@ 'k_on'
'b2-a3-10' B(b2), A(a3,a1!_,a2_ ) -> B(b2!0), A(a3!0,a1!_,a2_ )@ 'k_on'
'b2-a3-01' B(b2), A(a3,a1_ ,a2!_) -> B(b2!0), A(a3!0,a1_ ,a2!_)@ 'k_on'
'b2-a3-00' B(b2), A(a3,a1_ ,a2_ ) -> B(b2!0), A(a3!0,a1_ ,a2_ )@ 'k_on'

'b2 a3-11' B(b2!0), A(a3!0,a1!_,a2!_) -> B(b2), A(a3,a1!_,a2!_)@ 'k_off_tee'
'b2 a3-10' B(b2!0), A(a3!0,a1!_,a2_ ) -> B(b2), A(a3,a1!_,a2_ )@ 'k_off_vee'
'b2 a3-01' B(b2!0), A(a3!0,a1_ ,a2!_) -> B(b2), A(a3,a1_ ,a2!_)@ 'k_off_vee'
'b2 a3-00' B(b2!0), A(a3!0,a1_ ,a2_ ) -> B(b2), A(a3,a1_ ,a2_ )@ 'k_off'

```

B.3 Observables

Finally we need to define observables, so that the simulator understands what to plot. Here we observe **A**; **B** would be done similarly.

```
%var: 'A0' A(a1,a2,a3)
```

```

%var: 'a1' A(a1!_,a2,a3)
%var: 'a2' A(a1,a2!_,a3)
%var: 'a3' A(a1,a2,a3!_)

%var: 'A1' 'a1' + 'a2' + 'a3'

%var: 'a1a2' A(a1!_,a2!_,a3)
%var: 'a1a3' A(a1!_,a2,a3!_)
%var: 'a2a3' A(a1,a2!_,a3!_)

%var: 'A2' 'a1a2' + 'a1a3' + 'a2a3'

%var: 'A3' A(a1!_,a2!_,a3!_)

%plot: 'A0'
%plot: 'A1'
%plot: 'A2'
%plot: 'A3'

```

Finally, we observe 'alkenes' to get a sense of their frequency (very low as we have seen in the main text).

```

%var: 'Ba1a2' B(b1!1,b2!2),A(a1!1,a2!2)
%var: 'Ba2a1' B(b1!1,b2!2),A(a1!2,a2!1)
%var: 'Ba1a3' B(b1!1,b2!2),A(a1!1,a3!2)
%var: 'Ba3a1' B(b1!1,b2!2),A(a1!2,a3!1)
%var: 'Ba2a3' B(b1!1,b2!2),A(a2!1,a3!2)
%var: 'Ba3a2' B(b1!1,b2!2),A(a2!2,a3!1)

%var: 'alkene' 'Ba1a2'+ 'Ba2a1' + 'Ba1a3' + 'Ba3a1' + 'Ba2a3' + 'Ba3a2'

%plot: 'alkene'

```

This completes the definition of our example model. Each of the above modules can be written as a separate file which again increases clarity and flexibility.