The Annals of Statistics
2010, Vol. 38, No. 4, 2242–2281
DOI: 10.1214/09-AOS779

STOCHASTIC KINETIC MODELS: DYNAMIC INDEPENDENCE, MODULARITY AND GRAPHS[1]

BY CLIVE G. BOWSHER

University of Cambridge

The dynamic properties and independence structure of stochastic kinetic models (SKMs) are analyzed. An SKM is a highly multivariate jump process used to model chemical reaction networks, particularly those in biochemical and cellular systems. We identify SKM subprocesses with the corresponding counting processes and propose a directed, cyclic graph (the kinetic independence graph or KIG) that encodes the local independence structure of their conditional intensities. Given a partition $[A, D, B]$ of the vertices, the graphical separation $A \perp B|D$ in the undirected KIG has an intuitive chemical interpretation and implies that A is locally independent of B given $A \cup D$. It is proved that this separation also results in global independence of the internal histories of A and B conditional on a history of the jumps in D which, under conditions we derive, corresponds to the internal history of D. The results enable mathematical definition of a modularization of an SKM using its implied dynamics. Graphical decomposition methods are developed for the identification and efficient computation of nested modularizations. Application to an SKM of the red blood cell advances understanding of this biochemical system.

1. Introduction and summary. The dynamic properties and conditional independence structure of stochastic kinetic models are analyzed using a marked point process framework. A stochastic kinetic model or SKM is a highly multivariate jump process used to describe chemical reaction networks. SKMs have become particularly important as models of the network of interacting biomolecules in a cellular system. The necessity of a stochastic process approach to the dynamics of such biochemical reaction systems is now clear [28, 30], with SKMs providing continuous-time, mechanistic descriptions firmly grounded in chemical kinetic theory and the underlying statistical physics. The Gillespie algorithm [9, 10] for simulation of SKMs is now an important tool in the science of systems biology. However, there are few analytical tools for study of the dynamic properties of SKMs (although note [1, 12] and [8]), especially when the SKM is of modest or high dimension.

Received August 2009; revised December 2009.

[1]Supported by the EPSRC and MRC (United Kingdom).

AMS 2000 subject classifications. Primary 62P10, 62-09; secondary 60G55, 92C37, 92C40, 92C45.

Key words and phrases. Stochastic kinetic model, kinetic independence graph, counting and point processes, dynamic and local independence, graphical decomposition, reaction networks, systems biology.

This paper develops what appear to be the first methods for analyzing the local and global dynamic independence structure implied by a given SKM and shows how these may be used to uncover the modular architecture of the network at coarser or finer levels of resolution. The required information about the parameters of the SKM is modest, and consistent with the partial information about these currently available for many biochemical reaction networks. SKMs are often thought of as continuous-time, homogeneous Markov chains having nonfinite state space. However, the fact that there are a finite number of possible types of jump of the process—corresponding to the different types of possible biochemical reaction in the system—allows formulation of both the SKM and its subprocesses as multivariate counting processes. This turns out to be a fruitful approach for the problems addressed here. In fact, the Markov property is not needed for the results and methods of the paper. The main contributions may be summarized as follows.

Graphical models for SKMs and dynamic molecular networks are introduced. These kinetic independence graphs (KIGs) are directed, cyclic graphs whose vertices are the different types (or species) of biomolecule in the system. The KIG encodes local independences that result from a lack of dependence of the conditional intensity of a subprocess on the internal history of some of the species.

Given a partition $[A, B, D]$ of the vertices, the graphical separation $A \perp B|D$ in the undirected version of the KIG has an intuitive chemical interpretation and implies A is locally (or "instantaneously") independent of B given $A \cup D$ (and B locally independent of A given $B \cup D$). It is proved that this separation also results in conditional independence, over any finite time interval $(0, t]$, of the internal histories of A and B conditional on a history of the jumps in D. Conditions under which this history corresponds to the internal history of D are derived and are easily checked computationally. Such a conditional independence is termed a global (as opposed to local) dynamic independence here.

The new results enable mathematical definition of a modularization of an SKM using its implied dynamics. Graphical decomposition methods are developed for the identification of nested modularizations that allow the extent of coarse-graining to be varied and provide computationally efficient algorithms for large SKMs. Junction tree representations are shown to provide a useful tool for visualizing, summarizing and manipulating the modularizations. Applying the techniques of the paper to an SKM that represents detailed empirical knowledge of the metabolic network of the human red blood cell yields new insight into the biological organization and dynamics of this cellular system.

Graphical models and their associated analytical and computational methods allow the modularization of large, complex models into smaller components and provide a particularly effective means of representing and analyzing conditional independence relationships [3, 21]. Certain graphical approaches are now used quite extensively in computational biology and have also been readily assimilated by the

wider biological scientific community, which has long found diagrammatic representations of reaction schemes useful [15]. However, rigorous graphical representations of biochemical networks as dynamic processes—that is graphical models in the statistical sense—do not appear to have been considered previously.

Indeed, graphical models for continuous time stochastic processes in general are in an early stage of development. Didelez [5, 6] introduced graphs based on the local independence structure of conditional intensities for finite state, composable Markov processes and multivariate point processes, respectively; [22] is an earlier contribution, also for finite state Markov processes. SKMs require new methods since interest is in dynamic independences between groups of species rather than the counting processes for the different types of reaction per se. Furthermore, the Markov process for species concentrations implied by the SKM neither has finite state space, nor is it composable for most SKMs of interest (see Section 3).

In practice, the SKM is constructed from a large list of the biochemical reactions that comprise the network under study. This list, or "network reconstruction," is usually compiled using extensive experimental evidence in the literature on the component parts of the system and their molecular interactions [26]. Indeed, the approaches of molecular biology and genetics, including genome sequencing, have already proved remarkably successful in providing life scientists with a very extensive "parts list" for biology. Systems biology is an increasingly influential, interdisciplinary approach that aims to describe mathematically the stochastic dynamic behavior of the whole system as an emergent property of the network of interacting biomolecules [30].

A principal challenge is thus to map from fine level descriptions such as reaction lists and their implied SKMs to higher level, coarse-grained descriptions of the dynamic properties. Related is the increasingly held view that biochemical reaction networks are modular, that is their architecture can be decomposed into units that perform "nearly independently" [18], and that identifying such modules is a crucial step in the endeavor to understand and, ultimately, to selectively control cellular systems. However, it is recognized that rigorous, mathematical definition and identification of modularizations for biochemical networks is difficult, especially from a dynamic perspective (see [29], Chapter 3). As a result, such modularization techniques have been slow to develop, and there seems to be no prior work allowing for stochastic and non-steady state dynamics. The dynamic independence results and associated graphical methods developed here provide an effective means of addressing these problems. Broadly speaking, the paper also illustrates the utility of a statistical and probabilistic approach to the dynamics of biological systems which, despite their stochastic nature, have hitherto more often received the attention of physical scientists.

The structure of the paper is as follows. Section 2.1 introduces SKMs and reaction networks in a manner requiring no previous background in systems biology or biochemistry. Section 2.2 defines an SKM as a marked point process and provides a formal construction using the well-known Gillespie algorithm as a point of

departure. Section 2.3 then shows how to accommodate subprocesses of the SKM in a counting process framework and discusses their conditional intensities and internal histories (natural filtrations). Section 3 introduces the kinetic independence graphs, or KIGs, and examines local independence and graphical separation in the undirected KIG. Section 4 then relates these to global conditional independence of species histories in Theorems 4.4 and 4.5, which are central to the paper. Rigorous proofs of these theorems are quite involved and are given as Appendix A. Section 5 develops graphical decomposition methods and associated theory for the identification of modularizations of SKMs, while Section 6 applies the techniques of the paper to the SKM of the human red blood cell. Section 7 highlights some directions for future research.

2. SKMs and counting processes.

2.1. *Introducing the SKM and reaction networks*. A stochastic kinetic model is a continuous-time jump process modeling the state of a chemical system, $X(t) = [X_1(t), \ldots, X_n(t)]'$, where $X_i(t)$ is interpreted as the nonnegative, integer number of molecules of type i present at time t. The set of different types of molecule or the *species set* is given by $\mathcal{V} := \{1, \ldots, n\}$. There are a finite number of possible types of jump in $X(t)$ that may take place, corresponding to the different types of possible *reaction*, $m \in \mathcal{M} := \{1, \ldots, M\}$. It is particularly useful for our purposes to view an SKM as a marked point process or MPP in which the points or "events" correspond to the jump times of the process $X(t)$. Mathematically, a particular reaction can then be identified with an element of the finite mark space and each mark indicates the type of jump associated with the corresponding jump time.

An SKM is denoted here by $\{T_s, Z_s\}_{s\geq 1}$, where T_s is the sth jump time. The mark $Z_s \in \{S_m | m \in \mathcal{M}\}$ is the value of the jump and is interpreted as the *changes* in the number of molecules of each species. The matrix $S := [S_1, S_2, \ldots, S_M]$ is usually known as the stoichiometric matrix. Any two columns of S are taken to be nonequal; hence, there is a bijection between the mark space and $\mathcal{M}$. A formal construction of an SKM is given below in Section 2.2 but it is helpful at this stage to note the following linear equation determining the dynamic evolution of $X(t)$:

$$X(t) = X(0) + SN(t), \qquad t \geq 0,$$

where $N(t) = [N_1(t), \ldots, N_M(t)]'$ is the M-variate counting process associated with the marked point process $\{T_s, Z_s\}_{s\geq 1}$. Thus, $N_m(t)$ is interpreted as counting the number of reactions of type m during $(0, t]$. Denote by $\mathcal{F}_t^N := \sigma(N(s); 0 \leq s \leq t)$ the internal history of the entire process and by $\mathcal{F}_t^m := \sigma(N_m(s); 0 \leq s \leq t)$ the internal history of the mth counting process. The probability law of $N(t)$, and hence that of $X(t)$, is determined by what are known as the $\mathcal{F}_t^N$-conditional intensities, $[\lambda_m(t); m \in \mathcal{M}]$.

The conditional intensity concept is important for an understanding of the paper. At time t, each intensity $\lambda_m(t)$ is interpreted as the *local* (or *instantaneous*)

rate of reaction m, conditional on the internal history of *the entire process* $\mathcal{F}_t^N$. Confining attention to a finite interval of time $\mathcal{T}$, provided that $N(t)$ has finite expectation $\forall t \in \mathcal{T}$ (and that $[\lambda_m(t); t \in \mathcal{T}]$ is bounded by an integrable random variable), each intensity is a local rate of reaction in exactly the chemical sense—that is, $\lambda_m(t+) = \lim_{h \downarrow 0} \mathsf{E}[h^{-1}\{N_m(t+h) - N_m(t)\}|\mathcal{F}_t^N]$, the conditionally expected number of reactions of type m per unit time in the limit as h goes to zero. Of course, the intensities are themselves random variables (r.v.'s) since the evolution of N up to time t is itself a stochastic process, hence the appearance of the conditional expectation. A technical subtlety is that $\lambda_m(t)$ is defined to have sample paths that are left-continuous (with limits from the right), compared to the right-continuous sample paths of $X(t)$. A heuristic chemical interpretation is that if a jump in X takes place at t, then future jumps are (locally) determined by the intensity evaluated "immediately after" t.

A basic familiarity with the chemical representation and interpretation of reactions is helpful in what follows (see also [30] for an accessible introduction). Each reaction $m \in \mathcal{M}$ has the chemical representation

$$\sum_{i \in R[m]} \alpha_i X_i \to^m \sum_{j \in P[m]} \beta_j X_j, \tag{2.1}$$

which is read as follows: when reaction m takes place, α_i molecules of type i are consumed for each i in the subset $R[m] \subset \mathcal{V}$, and β_j molecules of type j are produced for each j in the subset $P[m] \subset \mathcal{V}$. The species $R[m]$ are called the *reactants* (or inputs) of the reaction m, and the species $P[m]$ are called the *products* (or outputs) of m. The integer coefficients $[\{\alpha_i\}, \{\beta_j\}]$ are known as the stoichiometries of the reaction. If a species k is a reactant but not a product, then its corresponding entry in the stoichiometric matrix S (i.e., the change in the level of k caused by reaction m) is given by $S_{km} = -\alpha_k$. Alternatively, if species k is a product but not a reactant, then $S_{km} = \beta_k$. There is no assumption that $R[m] \cap P[m] = \varnothing$, and if k is both a product and a reactant then $S_{km} = \beta_k - \alpha_k$. A common situation in this case is $\beta_k = \alpha_k$, that is k acts as a "catalyst," increasing the rate of the reaction but not itself being "changed" by the reaction—that is, not itself being overall consumed or produced when m takes place. Formally, the sets $R[m]$ and $P[m]$ are defined by allowing zero stoichiometries and writing the mth reaction as $\sum_{i \in \mathcal{V}} \alpha_i X_i \to^m \sum_{j \in \mathcal{V}} \beta_j X_j$. Then $R[m] := \{i \in \mathcal{V} | \alpha_i > 0\}$ and $P[m] := \{j \in \mathcal{V} | \beta_j > 0\}$.

In systems biology, a living cell is often viewed as a network of interacting biomolecules of different types, with n and M both large (and often $M > n$). The interaction is selective—only species that are reactants for some reaction m can together react to give products. Each reaction involves only a few species, so the cardinality of $R[m] \cup P[m]$ is small. Certain reactions are "coupled" in that a product of one reaction is also a reactant of another reaction. From a stochastic process perspective, the specification of the list of component reactions as in (2.1)

for all $m \in \mathcal{M}$ implies dependences between the levels (or concentrations) of the different biomolecules.

As a simple but nonetheless biochemically meaningful illustration, consider the following example of an SKM.

EXAMPLE 2.1. Consider the SKM with the 5 different species $\mathcal{V} = \{P, R, g, P_2, gP_2\}$ and the 6 reactions

$$g \to^{\text{trc}} g + R, \qquad R \to^{\text{trl}} R + P, \qquad 2P \to^{d} P_2,$$
$$P_2 \to^{rd} 2P, \qquad g + P_2 \to^{b} gP_2, \qquad gP_2 \to^{ub} g + P_2.$$

The gene (g) is responsible for the production of molecules of protein (P) via the intermediate (mRNA) species (R). In this simplified representation, g and R act as simple catalysts in the reactions trc ("transcription") and trl ("translation"), respectively. The third reaction d consists of the binding of 2 molecules of P (the sole reactant) to form the new molecule P_2 (the sole product). The fourth reaction rd is the reverse of the third. The fifth reaction sets up a "negative feedback cycle" whereby the production of P is negatively self-regulated by the binding of P_2 to g to form the distinct species gP_2. Genes bound in this way to P_2 are not then available to participate in the trc reaction, thus preventing over-production of the protein. We shall return later to the same example.

2.2. *Defining and constructing the SKM.* The Gillespie stochastic simulation algorithm [9, 10] has become an important tool in biological science for studying biochemical and cellular systems. Given its familiarity in mathematical and computational biology, the following construction of an SKM as a marked point process takes as its point of departure the conditional distributions employed in the Gillespie algorithm. For our purposes, the algorithm is usefully viewed as outputting a realization of the MPP $\{T_s, Z_s\}$, from which the resultant process $X(t)$ is easily constructed as in (2.4) below. Readers less concerned with formal constructions and already familiar with stochastic kinetics may proceed safely to Section 2.3 after noting Definition 2.1 of an SKM and (2.6) for the conditional reaction intensities (or "hazards").

Denote the numbers of molecules of all species at time T_s by $Z_s^X := Z_{s-1}^X + Z_s$ ($s = 1, 2, \ldots$). Let Z_0^X be the initial, deterministic state of the system, and define $T_0 := 0$. We write the σ-field generated by the first r points and marks as $\mathcal{F}_{T_s} := \sigma(T_r, Z_r; r = 1, \ldots, s)$. Also let $\mathcal{F}_{T_{s+1}-} := \sigma(T_{s+1}, T_r, Z_r; r = 1, \ldots, s)$, where the $(s+1)$th mark is excluded from the generating collection of random variables.

Now introduce the important propensity (or reaction rate) function for the mth reaction, $\lambda_m(Z_s^X)$, where $\lambda_m : \mathbb{N}_0^n \to [0, \infty)$ is continuous. The conditional distributions implied by stochastic kinetic theory [11] and employed in the Gillespie

algorithm are given by

$$\mathsf{P}(T_{s+1} > t | \mathcal{F}_{T_s}) = \exp\left\{-(t - T_s) \sum_{m=1}^{M} \lambda_m(Z_s^X)\right\}, \quad t > T_s, s = 0, 1, \ldots, \tag{2.2}$$

that is the waiting time to the next occurrence of a jump (reaction) is exponentially distributed with parameter $\sum_{m=1}^{M} \lambda_m(Z_s^X)$; and

$$\mathsf{P}(Z_{s+1} = S_m | \mathcal{F}_{T_{s+1}-}) = \lambda_m(Z_s^X) \Big/ \sum_{m=1}^{M} \lambda_m(Z_s^X), \tag{2.3}$$

which gives the mark (or jump) distribution. Note that both the waiting time and mark distributions depend only on Z_s^X, the levels of the species present following the sth reaction. The pure jump process $X(t)$ is given straightforwardly, for $t \geq 0$, by

$$X(t) := Z^X_{\max\{s\,:\,T_s \leq t\}}, \qquad X(0) := Z_0^X, \tag{2.4}$$

it being well known that $X(t)$ is a time-homogeneous Markov chain under P.

It turns out to offer significant advantages and simplification to adopt a MPP framework for the problems addressed in the paper. An SKM is thus defined here directly in terms of the MPP $\{T_s, Z_s\}$ and its corresponding counting processes. It is implicit in our definition of a MPP that $T_s < T_{s+1}$ whenever $T_s < \infty$ $(s \geq 1)$. Thus, reactions occur instantaneously and no two reactions ever have identical occurrence times in continuous time. The physical interpretation is that reaction durations are negligible and may be ignored. The random variables T_s are $(0, \infty]$-valued, with the interpretation that less than s reactions take place during the time interval $[0, \infty)$ if $T_s = \infty$. The flexibility gained will not be needed routinely, but may be useful for cellular systems that can enter an inactive or quiescent state. The stability condition $\lim_{s\to\infty} T_s = \infty$ a.s. is imposed, which is equivalent to the statement that only finitely many reactions occur in any finite time interval (sometimes known as nonexplosivity).

DEFINITION 2.1. A *stochastic kinetic model* (*SKM*) is the MPP $[\{T_s, Z_s\}_{s\geq 1}, S, \mathsf{P}]$ with mark space given by the columns of S, $\{S_m | m \in \mathcal{M}\}$, where no 2 columns of S are equal; and where the probability measure P is such that (2.2) holds P-a.s. on $\{T_s < \infty\}$, (2.3) holds P-a.s. on $\{T_{s+1} < \infty\}$, and $\lim_{s\to\infty} T_s = \infty$ P-a.s.

Equivalently, the SKM may be denoted by the corresponding multivariate counting process (MVCP), $[N, S, \mathsf{P}]$, where $N := [N_m(t); m \in \mathcal{M}]_{t\geq 0}$, and $N_m(t) = \sum_{s\geq 1} 1(T_s \leq t) 1(Z_s = S_m)$ counts the number of reactions of type m that occur during $[0, t]$.

Note that by definition the reaction counting processes $\{N_m(t); m \in \mathcal{M}\}$ have no jump times in common. If the stability condition $T_s \to \infty$ a.s. holds, there exists for any propensity functions $\lambda_m(X_{t-})$—see [16], Theorem 1.7, page 56—a unique or *canonical* SKM satisfying Definition 2.1 on $(\Omega, \mathcal{F})$, where Ω is the space of M-variate counting process paths ([16], Definition 1.2, page 53), N is the identity map from $\Omega \to \Omega$, and $\mathcal{F} = \sigma(N(t); t \geq 0)$.

It follows from (2.2) and (2.3) that the propensity functions give the $\mathcal{F}_t^N$-conditional intensity process $\lambda(t)$ in the MVCP sense (see [16], Definition 2.7), that is, $\lambda(t) = [\lambda_m(X_{t-})]_{m\in\mathcal{M}}$. When $N(t)$ has finite expectation $\forall t > 0$, this means that $[N_m(t) - \int_0^t \lambda_m(s)\,ds]$ is an $\mathcal{F}_t^N$-martingale $\forall m$. That the intensities satisfy

$$\begin{aligned} \lim_{h\downarrow 0} \frac{1}{h} \mathsf{P}\big(N_m(t+h) - N_m(t) = 1 | \mathcal{F}_t^N\big) &= \lambda_m(X_t), \qquad m \in \mathcal{M}, \\ \lim_{h\downarrow 0} \frac{1}{h} \mathsf{P}\big(\bar{N}(t+h) - \bar{N}(t) > 1 | \mathcal{F}_t^N\big) &= 0, \end{aligned} \tag{2.5}$$

where $\bar{N}(t) := \sum_{m\in\mathcal{M}} N_m(t)$, is in fact a principle conclusion of the arguments of stochastic kinetic theory [11]. The assumptions of the theory are that the system is spatially homogeneous (or "well-stirred"), confined to a fixed volume and held at constant temperature. Under these assumptions, (2.2) and (2.3) have a firm physico-chemical basis [11].

It plays a significant role in what follows that the theory implies that the $\mathcal{F}_t^N$-intensities, $\lambda_m(t)$, have the form

$$\lambda_m(t) = c_m g_m\big\{X^{R[m]}(t-)\big\}, \tag{2.6}$$

where $c_m > 0$ is a deterministic ("rate") constant, and $g_m\{\cdot\} \geq 0$ is a continuous function depending *only on the levels of the reactants* $R[m]$.

2.3. *SKM subprocesses—histories and intensities.* For any subset of molecular species $A \subseteq \mathcal{V}$, let the vector process $\{X^A(t)\} := \{X_i(t); i \in A\}$ denote the corresponding subprocess of X. We identify X^A with its MVCP, analogously to the treatment of $X = X^{\mathcal{V}}$ above. For $A \subseteq \mathcal{V}$, consider the subset of reactions $\Delta(A) \subseteq \mathcal{M}$ that change (the level of) A, that is, $\Delta(A) := \{m \in \mathcal{M} : S_m^A \neq \mathbf{0}\}$, where S_m^A is the subvector of S_m corresponding to the elements of A. One can identify X^A with the MPP, $\{T_s^A, Z_s^A\}$, where each jump time T_s^A corresponds to the occurrence of some reaction in $\Delta(A)$; the mark Z_s^A gives the resultant jumps in the elements of A and takes its value in the mark space $E^A := \{S_m^A | m \in \Delta(A)\}$. This results in the following definition of an SKM subprocess.

DEFINITION 2.2. Let $[N, S, \mathsf{P}]$ be an SKM and for $A \subset \mathcal{V}$, let $\Delta(A)$ be the nonempty, finite subset $\{m \in \mathcal{M} : S_m^A \neq \mathbf{0}\}$. Denote by $\mathcal{M}(\Delta(A))$ the partition of

$\Delta(A)$ obtained by grouping reactions that change A identically, that is by applying to $\Delta(A)$ the equivalence relation

$$m \sim_A m' \quad \Leftrightarrow \quad S_m^A = S_{m'}^A.$$

Denote the eth element of $\mathcal{M}(\Delta(A))$ by $\mathcal{M}_e(\Delta(A))$, $e = 1, \ldots, |\mathcal{M}(\Delta(A))|$. The *subprocess of the SKM*, $N^A(t)$, is the $|\mathcal{M}(\Delta(A))|$-variate counting process given by

$$N^A(t) := \left\{ \sum_{m \in \mathcal{M}_e(\Delta(A))} N_m(t) \right\}_{e=1,\ldots,|\mathcal{M}(\Delta(A))|}. \tag{2.7}$$

The internal history of $N^A(t)$ is denoted $\mathcal{F}_t^A$.

Note that since $\mathcal{M}(\Delta(A))$ is a partition of $\Delta(A)$, the components of $N^A(t)$ have no jumps in common. Each element of the MVCP $N^A(t)$ thus counts the number of times reactions in $\Delta(A)$ have occurred that result in a given change in A. Intuitively, putting these elements together for all possible types of change in A to form a sample path of $N^A(t)$ captures exactly the "information" given by the corresponding sample path of $X^A(t)$. Indeed, there is a bijection between the sample paths of $N^A(t)$ and those of $X^A(t)$. The following technical lemma establishes that the internal history of the MVCP $N^A(t)$ is identical to that of $X^A(t)$.

LEMMA 2.1. *For $A \subseteq \mathcal{V}$, let $N^A(t)$ be a subprocess of an SKM as in Definition 2.2 and let $\mathcal{F}_t^{X_A} := \sigma(X^A(s); s \le t)$ be the internal history of the jump process $X^A(t)$. Then $\mathcal{F}_t^A = \mathcal{F}_t^{X^A}$ $\forall t \ge 0$. Furthermore, if $[A, B, D]$ is a partition of $\mathcal{V}$ then $\mathcal{F}_t^N = \mathcal{F}_t^A \vee \mathcal{F}_t^B \vee \mathcal{F}_t^D = \mathcal{F}_t^X$, $\forall t \ge 0$.*

PROOF. A proof that $\mathcal{F}_t^A = \mathcal{F}_t^{X^A}$ is given in Appendix B. For $\mathcal{F}_t^N = \mathcal{F}_t^X$, take $A = \mathcal{V}$. Finally, $\mathcal{F}_t^X = \mathcal{F}_t^{X^A} \vee \mathcal{F}_t^{X^B} \vee \mathcal{F}_t^{X^D}$ since $X(t) = [X^A(t)', X^B(t)', X^D(t)']'$. □

One advantage of a counting process definition of the subprocess for the species in $A \subset \mathcal{V}$ is that one may speak of the $\mathcal{F}_t^N$-intensity for the subprocess and interpret this in the usual manner as determining the *local* or *instantaneous* dependence of the subprocess on the full internal history of the SKM, $\mathcal{F}_t^N$.

PROPOSITION 2.2. *For $A \subseteq \mathcal{V}$, let $N^A(t)$ be a subprocess of an SKM as in Definition 2.2. The $\mathcal{F}_t^N$-conditional intensity under* P *is given by*

$$\lambda^A(t) := \left\{ \sum_{m \in \mathcal{M}_e(\Delta(A))} \lambda_m(t) \right\}_{e=1,\ldots,|\mathcal{M}(\Delta(A))|}. \tag{2.8}$$

PROOF. Immediate from (2.7) on noting that the intensities of the superpositions of the counting processes are the sums of the corresponding intensities. □

Notice that each element of the intensity, $\lambda_e^A(t)$, is the sum of the intensities (or stochastic rates) of all those reactions that result in the corresponding change in A.

It follows from the equations in (2.5) that, for any $A \subseteq \mathcal{V}$, the probability conditional on $\mathcal{F}_t^N$ that, during $(t, t+h]$, there is no change in X^A is equal to $1 - h\sum_{e=1}^{|\mathcal{M}(\Delta(A))|} \lambda_e^A(t) + o(h)$. Similarly, the probability conditional on $\mathcal{F}_t^N$ that, during $(t, t+h]$, there is exactly one jump in X^A equal to S_m^A, for some $m \in \mathcal{M}_e(\Delta(A))$, and also that no other reaction $m' \in \mathcal{M}$ occurs is equal to $h\lambda_e^A(t) + o(h)$ for $e = 1, \ldots, |\mathcal{M}(\Delta(A))|$. Summing over all of the foregoing, mutually exclusive events shows that these have conditional probability equal to $1 + o(h)$. Thus, in this infinitesimal sense, the $\mathcal{F}_t^N$-intensity $\lambda^A(t)$ may be interpreted as determining the local dependence of $N^A(t)$ on $\mathcal{F}_t^N$.

3. Kinetic independence graphs. The identification of subprocesses of the SKM with their corresponding MVCPs (see Section 2.3) greatly facilitates the construction of a *kinetic independence graph* encoding the local independence structure of the SKM—see Definition 3.1 below. The use of the local independence concept in constructing graphical models for continuous time processes owes much to Didelez [5, 6]. However, SKMs require new methods since interest is in dynamic independences between groups of species rather than the reaction counting processes $[N_m(t)]_{m\in\mathcal{M}}$ per se. Thus, the vertex set of the graph will be $\mathcal{V}$ rather than $\mathcal{M}$.

It is worth noting that existing graphical models for continuous-time Markov chains [5, 22] are not applicable to SKMs because the Markov process $X(t)$ neither has finite state space, nor is it composable for most SKMs of interest. Roughly speaking, composability [5] implies that any change of state in $X(t)$ can be represented as a change in only one of several components. Consider the use of $X^A(t)$ and $X^{\mathcal{V}\setminus A}(t)$ as components [since if $X(t)$ is composable with more than 2 subsets of species as components, it must be composable with just 2 components]—either the paths of $X^A(t)$ and $X^{\mathcal{V}\setminus A}(t)$ have common jump times contradicting that $X(t)$ is composable, or they constitute 2 separate SKMs which then require a new method for their individual analysis.

The kinetic independence graph of an SKM is defined as follows.

DEFINITION 3.1. The directed graph G with vertex set $\mathcal{V}$ is the *kinetic independence graph* (*KIG*) of the SKM $[N, S, \mathsf{P}]$ if and only if

$$\mathrm{pa}(k) = R[\Delta(k)] \setminus \{k\} \qquad \forall k \in \mathcal{V}, \tag{3.1}$$

where $\mathrm{pa}(k) = \{i \in \mathcal{V} | i \to k\}$ is the set of parents of vertex k, and $R[\Delta(k)] := \bigcup_{m\in\Delta(k)} R[m]$ is the set of reactants of all reactions that change species k.

Since only partial information about the SKM is required for construction of the KIG, the necessary information is currently available for many biochemical reaction networks. For each $m \in \mathcal{M}$, it is required to know the reactants $R[m]$, and the species (reactants and products) changed by the reaction, that is, $\{i \in \mathcal{V} | S_{im} \neq 0\}$. Full knowledge of the stoichiometric matrix S is neither necessary nor sufficient for construction of the KIG. Note that the possible presence of a catalyst among the reactants $R[m]$ implies that $R[m]$ cannot be reliably reconstructed from S. No knowledge of the rate parameters c_m is required for construction of the KIG, which is important since their measurement is difficult experimentally.

A comment will be useful at this juncture on the treatment of measurability considerations in the paper. While the treatment is fully rigorous, it is appreciated that some readers will be more concerned with application of the paper's results. Proofs requiring a measure-theoretic approach have therefore been placed in Appendices A and B. Note that a statement such as the one that $\lambda_m(t)$ is (as it must be) measurable $\mathcal{F}_t^{R[m]}$ implies that the realized value of the r.v. $\lambda_m(t)$ may be "computed" from the sample path of the subprocess for $R[m]$ over the interval $[0, t]$.

The motivation for Definition 3.1 of the KIG of an SKM is that the local evolution of species k depends only on the stochastic rate of reactions that change the number of molecules (the level) of k, which in turn depend only on the levels of their reactants. To make this exact, the concept of local independence [6] is needed. Let $A, B \subset \mathcal{V}$. We will say that N^B is *locally independent* of N^A (given $N^{\mathcal{V}\setminus A}$) if and only if the $\mathcal{F}_t^N$-intensity, $\lambda^B(t)$, is measurable $\mathcal{F}_t^{\mathcal{V}\setminus A}$ for all t—that is, the internal history of X_t^A is irrelevant for the $\mathcal{F}_t^N$-intensity of the species in B. Only intensities of subprocesses conditional on the history of the whole system, $\mathcal{F}_t^N$, are considered here (as opposed to $\mathcal{G}_t$-intensities where $\mathcal{G}_t \subset \mathcal{F}_t^N$).

As a consequence of Definition 3.1, one can read off from the KIG, for any collection of vertices B, those subprocesses with respect to which N^B is locally independent, that is, which are irrelevant for the instantaneous evolution of B. Denote the closure of B by $\mathrm{cl}(B) := \mathrm{pa}(B) \cup B$.

PROPOSITION 3.1. *Let G be the KIG of the SKM $[N, S, \mathsf{P}]$ and let $A, B \subset \mathcal{V}$. Then the $\mathcal{F}_t^N$-intensity $\lambda^B(t)$ is measurable $\mathcal{F}_t^{\mathrm{cl}(B)}$ for all t, that is, N^B is locally independent of $N^{\mathcal{V}\setminus\mathrm{cl}(B)}$ (given $N^{\mathrm{cl}(B)}$). Suppose that $A \cap \mathrm{cl}(B) = \varnothing$. Then $\lambda^B(t)$ is measurable $\mathcal{F}_t^{V\setminus A}$.*

PROOF. By (2.8), each intensity $\lambda_e^B(t)$ is measurable $\mathcal{F}_t^{R[\Delta(B)]}$ because $\lambda_m(t) = c_m g_m\{X^{R[m]}(t-)\}$ is measurable $\mathcal{F}_t^{X^{R[\Delta(B)]}} = \mathcal{F}_t^{R[\Delta(B)]}$ $\forall m \in \Delta(B)$, recalling that $R[\Delta(B)] = \bigcup_{m\in\Delta(B)} R[m]$. Since $\mathrm{pa}(B) = \{\bigcup_{k\in B} \mathrm{pa}(k)\} \setminus \{B\} = \{\bigcup_{k\in B} R[\Delta(k)]\} \setminus \{B\} = \{R[\Delta(B)]\} \setminus \{B\}$, it follows that $R[\Delta(B)] \subseteq \mathrm{cl}(B)$, and hence $\mathcal{F}_t^{R[\Delta(B)]} \subseteq \mathcal{F}_t^{\mathrm{cl}(B)}$ by Lemma 2.1. Thus, each intensity $\lambda_e^B(t)$ is measurable $\mathcal{F}_t^{\mathrm{cl}(B)}$, and the remainder of the proposition follows immediately. □

Proposition 3.1 accords with chemical intuition. Given the internal history of $X^{\mathrm{cl}(B)}$ at time t, the levels of the species $R[\Delta(B)]$ just prior to t are "known." These are exactly the species levels that determine the local dynamics of B since, as reactants, they determine the rate of all reactions that change the concentrations of B. Therefore, any further information about species histories, including the internal history of $N^{\mathcal{V}\setminus\mathrm{cl}(B)}$, is irrelevant for the local dynamics of B.

Notice that loops, that is edges of the type $k \to k$ are by definition not included in the KIG, even though k may well be in $R[\Delta(k)]$. For this reason, one cannot assert in Proposition 3.1 that $\lambda^B(t)$ is measurable $\mathcal{F}_t^{\mathrm{pa}(B)}$, but rather that it is measurable $\mathcal{F}_t^{\mathrm{cl}(B)}$. More generally, a particular SKM may imply further local independences of $\lambda^B(t)$ than those encoded by the KIG—for example, due to a deterministic relationship between two subsets of species arising from a chemical conservation relation—but this level of knowledge about the SKM is not assumed in constructing the KIG.

Graphical separations in the undirected version of the KIG, written $G^{\sim}$, are central in what follows. Diagrammatically, $G^{\sim}$ is the undirected graph obtained from G by substituting lines for arrows. Let $A, B, D \subset \mathcal{V}$. The notation $A \perp_{G^{\sim}} B|D$ stands for the *graphical separation* of A from B by D, that is, the property that every sequence of edges (or path) in $G^{\sim}$ that begins with some $a \in A$ and (without any repetition of vertices) ends with some $b \in B$, includes a vertex in D. With $[A, B, D]$ a partition of $\mathcal{V}$, such a separation in $G^{\sim}$ is equivalent to the nonexistence of $(a \in A, b \in B)$ such that there is an edge $a \to b$ or an edge $b \to a$ in G. This graphical separation implies the following mutual local independence property.

PROPOSITION 3.2. *Let G be the KIG of an SKM $[N, S, \mathsf{P}]$, and let $[A, B, D]$ be a partition of $\mathcal{V}$. If $A \perp_{G^{\sim}} B|D$, then N^B is locally independent of N^A (given $N^{B\cup D}$) and N^A is locally independent of N^B (given $N^{A\cup D}$) or, equivalently, $\lambda^B(t)$ is measurable $\mathcal{F}_t^{B\cup D}$ and $\lambda^A(t)$ is measurable $\mathcal{F}_t^{A\cup D}$.*

PROOF. $A \perp_{G^{\sim}} B|D$, if and only if $B \cap \mathrm{cl}(A) = A \cap \mathrm{cl}(B) = \varnothing$. The result then follows directly from Proposition 3.1. □

Note that it follows from the definition of the KIG that the graphical separation in Proposition 3.2 is equivalent to the chemical property $A \cap R[\Delta(B)] = B \cap R[\Delta(A)] = \varnothing$. That is, A does not participate as a reactant in any reaction that changes B, and vice versa. Therefore, for example, $R[\Delta(B)] \subseteq B \cup D$—hence, given the levels of B and D (which fully determine the rate of reactions that change B), the levels of A are irrelevant for the instantaneous evolution of B, and vice versa. Section 4 will establish that under weak regularity conditions on the SKM, the separation $A \perp_{G^{\sim}} B|D$ in $G^{\sim}$ implies not only mutual local independence but also *global* conditional independence of the internal histories of A and B given a history of D.

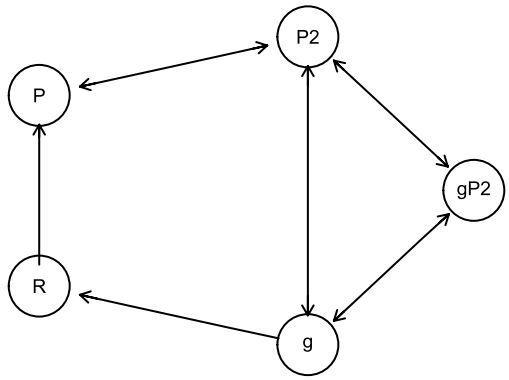

FIG. 1. *Kinetic independence graph of the SKM in Example 2.1.*

As an illustration of the concepts discussed so far, consider again the SKM of Example 2.1. The corresponding KIG is shown in Figure 1. Note the presence of cycles in the KIG, including $g \to R \to P \to P_2 \to g$ which might be termed the "negative feedback cycle." Clearly, $\{P, R\} \perp_{G^\sim} \{gP_2\}|\{g, P_2\}$. Let $D := \{g, P_2\}$. Notice that, according to Definition 2.2, SKM subprocesses are given by $N^{gP_2} = [N_b, N_{ub}]'$ and $N^D = [N_d, N_{rd}, N_b, N_{ub}]'$. Hence, $\mathcal{F}_t^{gP_2} \subset \mathcal{F}_t^D$, and the global independence $\mathcal{F}_t^{P,R} \perp\!\!\!\perp \mathcal{F}_t^{gP_2}|\mathcal{F}_t^D$ holds immediately in this case.

Anticipating the problem of how to modularize SKMs to be tackled in Section 5, a modularization suggested for the SKM of Example 2.1 by its dynamic independence properties is the modularization $[\{P, R, g, P_2\}, \{gP_2, g, P_2\}]$. The 2 module "residuals" are given by $\{P, R\}$ and $\{gP_2\}$. Each module residual is locally independent of the other given that module's internal history. Furthermore, the 2 modules are conditionally independent given the history of their intersection, $\mathcal{F}_t^{g,P_2}$. In fact, these 2 modules correspond to the maximal prime subgraphs of $G^\sim$ for this example (see Definition 5.2). The graphical methods for identifying SKM modularizations in Section 5 are, broadly speaking, also based around the maximal prime decomposition of the undirected KIG.

4. Global dynamic independence. This section will present the theorems establishing that for a partition $[A, B, D]$ of $\mathcal{V}$, the separation of A from B by D in the undirected version of the KIG implies the global dynamic independence $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B|\mathcal{F}_t^{D^*}$ for all $t \geq 0$, under the probability measure of the SKM, P.

The history of X^D given by $\mathcal{F}_t^{D^*}$ is defined formally below. Heuristically, $\mathcal{F}_t^{D^*}$ includes at time t the internal history of the jump process for the species in D ($\mathcal{F}_t^D \subseteq \mathcal{F}_t^{D^*}$), and also for every jump time of X^D always contains the "information" whether some species in A jumped, or some species in B jumped, or some species in both A and B jumped, but not (necessarily) the particular species involved. The main proofs of the theorems are quite involved and are given in Appendix A. Readers less concerned with technical details will find an outline of the

argument and intuitions for important aspects of the proofs in this section of the paper. It is worth defining here explicitly what is meant by the conditional independence of σ-fields [4, 7].

DEFINITION 4.1. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be an arbitrary probability space and suppose we have 3 sub-σ-fields $\mathcal{F}^1, \mathcal{F}^2, \mathcal{F}^3 \subseteq \mathcal{F}$. We say that $\mathcal{F}^1$ and $\mathcal{F}^2$are independent conditionally on $\mathcal{F}^3$ and write $\mathcal{F}^1 \perp\!\!\!\perp \mathcal{F}^2 | \mathcal{F}^3; \mathsf{P}$ if and only if

$$\mathsf{E}[Z_1 | \mathcal{F}^2 \vee \mathcal{F}^3] = \mathsf{E}[Z_1 | \mathcal{F}^3]$$

for all nonnegative random variables Z_1 that are measurable $\mathcal{F}^1$. The notation $\mathcal{F}^2 \vee \mathcal{F}^3$ stands for the smallest σ-field containing both $\mathcal{F}^2$ and $\mathcal{F}^3$. The relationship is symmetric, that is, $\mathcal{F}^1 \perp\!\!\!\perp \mathcal{F}^2 | \mathcal{F}^3; \mathsf{P} \Leftrightarrow \mathcal{F}^2 \perp\!\!\!\perp \mathcal{F}^1 | \mathcal{F}^3; \mathsf{P}$.

Thus, the global dynamic independence statement $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B | \mathcal{F}_t^{D^*}$ can be understood as follows: the expectation of (suitably measurable) mappings from sample paths of N^A (resp., N^B) over $(0, t]$ to $\mathbb{R}$, conditional on the history $\mathcal{F}_t^{D^*}$, are unchanged when the conditioning σ-field also includes the internal history of N^B (resp., N^A). Roughly speaking, and over any time interval $(0, t]$, all "information" about the dynamic evolution of B is irrelevant for the dynamic evolution of A, given the "information" in $\mathcal{F}_t^{D^*}$ (and vice versa).

First, an outline of the logic of the argument of this section is presented, before going on to state the main theorems.

4.1. *Preliminaries and outline of argument.* The following lemma is central to the method used. Although closely related to a result in [7], I am not aware of its statement and proof elsewhere.

LEMMA 4.1. *Let* $\mathsf{P}, \tilde{\mathsf{P}}$ *be probability measures on an arbitrary measurable space,* $(\Omega, \mathcal{F})$, *such that* $\mathsf{P} \ll \tilde{\mathsf{P}}$. *Consider any* 3 *sub-*σ*-fields* $\mathcal{F}^1, \mathcal{F}^2, \mathcal{F}^3 \subseteq \mathcal{F}$ *satisfying the conditional independence* $\mathcal{F}^1 \perp\!\!\!\perp \mathcal{F}^2 | \mathcal{F}^3; \tilde{\mathsf{P}}$ *under the dominating measure* $\tilde{\mathsf{P}}$. *Denote by* $\mathcal{L}_{123}$ *a Radon–Nikodym derivative,* $(\mathsf{dP}_t / \mathsf{d}\tilde{\mathsf{P}}_t)|_{\mathcal{F}^1 \vee \mathcal{F}^2 \vee \mathcal{F}^3}$.

Then the following condition implies that the conditional independence $\mathcal{F}^1 \perp\!\!\!\perp \mathcal{F}^2 | \mathcal{F}^3$ *holds also under* P:

$$\mathcal{L}_{123} = \psi_{13} \psi_{23},$$

where ψ_{i3} *is a nonnegative,* $\mathcal{F}^i \vee \mathcal{F}^3$*-measurable random variable for* $i \in \{1, 2\}$.

Proof of Lemma 4.1 is given in Appendix B.

Since $\mathcal{F}_t^{D^*}$ is a history of N^D (i.e., $\mathcal{F}_t^D \subseteq \mathcal{F}_t^{D^*}$), it follows from Lemma 2.1 that $\mathcal{F}_t^A \vee \mathcal{F}_t^B \vee \mathcal{F}_t^{D^*} = \mathcal{F}_t^X = \mathcal{F}_t^N$. A likelihood process $\mathcal{L}_t := (\mathsf{dP}_t / \mathsf{d}\tilde{\mathsf{P}}_t)|_{\mathcal{F}_t^N}$ is thus required in order to apply Lemma 4.1 to the 3 σ-fields $\mathcal{F}_t^A, \mathcal{F}_t^B, \mathcal{F}_t^{D^*}$. Given its importance here, we restate for an SKM the following likelihood result from the counting process literature (see, e.g., [16], Theorem 4.1, page 74). The proof is omitted since it is well known.

LEMMA 4.2. *Let* $[N, S, \mathsf{P}]$ *be an SKM as in Definition* 2.1, *and let* $[N, \tilde{\mathsf{P}}]$ *be the* M*-variate Poisson process with intensities* $\mathbf{1}_M = (1, \ldots, 1)'$. *Then, for every* $t \geq 0$, $\mathsf{P}_t \ll \tilde{\mathsf{P}}_t$ *and a Radon–Nikodym derivative is given by*

$$\left.\frac{d\mathsf{P}_t}{d\tilde{\mathsf{P}}_t}\right|_{\mathcal{F}_t^N} = \prod_{m=1}^{M} \left\{ \prod_{T_s^m \leq t} \lambda_m(X_{T_s^m -}) \right\} \exp\left\{ t - \int_0^t \lambda_m(X_{u-})\, du \right\}. \tag{4.1}$$

Note that the counting processes $[N_m; m = 1, \ldots, M]$ *are independent under* $\tilde{\mathsf{P}}$ (*see, e.g.*, [17], *Proposition* 4.7.2), *and hence the* σ*-fields* $[\mathcal{F}_t^m; m = 1, \ldots, M]$ *are independent under* $\tilde{\mathsf{P}}$ *and* $\tilde{\mathsf{P}}_t$ *for all* $t \geq 0$.

Of considerable importance here will be the fact that, under the dominating measure $\tilde{\mathsf{P}}$, the counting processes $[N_m; m = 1, \ldots, M]$ are independent. Of course, two or more of the subprocesses $[N^A, N^B, N^D]$ may have jump times in common as the result of reactions that simultaneously change several of the species sets $[A, B, D]$. However, denoting by ΔD_D the reactions that change D alone, the reaction set $\mathcal{M}$ can be partitioned as $[\Delta(A), \Delta(B) \setminus \Delta(A), \Delta D_D]$. The independence of the reaction counting processes then implies that $\mathcal{F}_t^{\Delta(A)} \perp\!\!\!\perp \mathcal{F}_t^{\Delta(B)\setminus\Delta(A)} | \mathcal{F}_t^{\Delta D_D}; \tilde{\mathsf{P}}_t$, which is a point of departure for proving Theorem 4.4 below.

To apply Lemma 4.1, we first establish that if $A \perp_{G^\sim} B | D$, then $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B | \mathcal{F}_t^{D^*}$ under the dominating measure $\tilde{\mathsf{P}}$ (see Theorem 4.4). We then show that the factorisation $\mathcal{L}_t = \psi_{AD^*,t}\psi_{BD^*,t}$ holds with, for example, $\psi_{AD^*,t}$ an $\mathcal{F}_t^A \vee \mathcal{F}_t^{D^*}$-measurable r.v. (see Theorem 4.5). We are thus able to conclude that if $A \perp_{G^\sim} B | D$, then the SKM must satisfy $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B | \mathcal{F}_t^{D*}; \mathsf{P}$ (see Corollary 4.6).

Definition of the filtration $\{\mathcal{F}_t^{D^*}\}$ is needed. This is best understood as the internal history of a particular MVCP, $N^{D^*}(t)$, defined now below.

DEFINITION 4.2. Let $[A, B, D]$ be a partition of $\mathcal{V}$, the species set of an SKM. Define $\Delta D_A := \{\Delta(D) \cap \Delta(A)\} \setminus \Delta(B)$, the set of reactions that change D and A, but not B. Similarly, define $\Delta D_{AB} := \Delta(D) \cap \Delta(A) \cap \Delta(B)$, $\Delta D_B := \{\Delta(D) \cap \Delta(B)\} \setminus \Delta(A)$, and those that change D alone by $\Delta D_D := \Delta(D) \setminus \{\Delta(A) \cup \Delta(B)\}$. Then $[\Delta D_A, \Delta D_{AB}, \Delta D_B, \Delta D_D]$ is a partition of $\Delta(D)$, the reactions that change D.

The MVCP $N^{D^*}(t)$ is constructed by taking each element, $\Delta D_\bullet$, of this partition in turn and summing over counting processes for reactions in $\Delta D_\bullet$ alone that result in identical changes to D—that is, by applying to each $\Delta D_\bullet$ the equivalence relation

$$m \sim m' \quad \Leftrightarrow \quad S_m^D = S_{m'}^D, \qquad m, m' \in \Delta D_\bullet.$$

The resultant MVCP, denoted $N_\bullet^D(t)$, is given by

$$N_\bullet^D(t) := \left\{ \sum_{m \in \mathcal{M}_e(\Delta D_\bullet)} N_m(t) \right\}_{e=1,\ldots,|\mathcal{M}(\Delta D_\bullet)|},$$

which differs from the SKM subprocess $N^D(t)$ of Definition 2.2 in that the partition $\mathcal{M}(\Delta D_\bullet)$ is used in place of $\mathcal{M}(\Delta(D))$. Then define

$$N^{D^*}(t) := [N_A^D(t), N_{AB}^D(t), N_B^D(t), N_D^D(t)].$$

Here $\Delta D_\bullet$ is empty, the relevant component of $N^{D^*}(t)$ is set equal to zero $\forall t \geq 0$. The corresponding internal histories of $N_\bullet^D(t)$ are written $\{\mathcal{F}_A^D(t), \mathcal{F}_{AB}^D(t), \mathcal{F}_B^D(t), \mathcal{F}_D^D(t)\}$.

Thus, as stated previously, at time t $\mathcal{F}_t^{D^*}$ includes the internal history of the jump process for the species in D, $\mathcal{F}_t^D$, and also for every jump time of X^D always contains the "information" whether some species in A jumped, or some species in B jumped, or some species in both A and B jumped, but not (necessarily) the particular species involved. An alternative formulation of $\mathcal{F}_t^{D^*}$ would be as the internal history at t of the marked point process $\{T_s^D, \tilde{Z}_s^D\}$, where the marks $\tilde{Z}_s^D$ give not only the value of the jump in D but also an indicator of which element, $\Delta D_\bullet$, the reaction causing the jump in D belongs to.

In some applications, it may be more convenient or practical to use only internal histories. Section 4.3 will thus provide a rigorous and intuitive means of comparing $N^{D^*}(t)$ and $N^D(t)$—through comparison of the corresponding partitions of the reactions in $\Delta(D)$—and state a property that is easily checked for a given SKM under which $\mathcal{F}_t^{D^*} = \mathcal{F}_t^D$.

The following conditions on the SKM are used in the statement of the results of this section. Both Theorems 4.4 and 4.5 assume that the SKM is standard, which imposes the following very weak regularity conditions.

DEFINITION 4.3. An SKM $[N, S, \mathsf{P}]$ is a *standard SKM* if it satisfies all of the following: (i) every reaction changes at least 1 species, that is, $S_m \neq \mathbf{0}$ $\forall m \in \{1, \ldots, M\}$; (ii) every species in $\mathcal{V}$ is changed by at least one reaction, that is, the row $S_{k\bullet} \neq \mathbf{0}$ $\forall k \in \mathcal{V}$; (iii) if a zeroth order reaction $\tilde{m}$ is included (i.e., $R[\tilde{m}] = \varnothing$) then it has only 1 product; (iv) for all m, if $|R[m]| = 1$ then $|R^*[m]| = 1$ and if $|R[m]| > 1$ then $|\{R[m]\} \setminus \{R^*[m]\}| \leq 1$, where $R^*[m] = \{i \in R[m] | S_{im} \neq 0\}$ are the reactants changed by m.

The first condition of Definition 4.3 is obvious. The second does not preclude an effect of the concentration of species that are constant over time (via the functions g_m or the constants c_m). The third is just a convention. The fourth ensures that if $R[m] \neq \varnothing$ then the reaction has at least 1 reactant that is changed and at most 1 reactant that is not changed. It allows for the inclusion of a reaction with an unchanged reactant where this simplifies the SKM, for example, where that reactant acts as a catalyst [although the reaction could be broken down into several reactions not requiring condition (iv) if desired].

Theorems 4.4 and 4.5 also require that the following condition holds for $\Gamma = \Delta(A) \cap \Delta(B)$. If $\Delta(A) \cap \Delta(B) = \varnothing$, as sometimes happens, then the condition is trivial and always satisfied.

CONDITION 4.3. Let $[N, S, \mathsf{P}]$ be a standard SKM. A subset of reactions Γ, $\varnothing \subseteq \Gamma \subseteq \mathcal{M}$, is said to be identified by consumption of reactants if and only if:
(i) For all $m \in \Gamma$, $S_{im} \leq 0$ $\forall i \in R[m]$ (hence, $S_{km} < 0$ for some $k \in R[m]$ provided that $R[m] \neq \varnothing$), and $S_{im} \geq 0$ $\forall i \in P[m]$; and (ii) $\nexists m, \tilde{m} \in \Gamma$ $(m \neq \tilde{m})$ such that $S_m^- = S_{\tilde{m}}^-$, where S_m^- denotes the vector formed by setting all positive elements of S_m to zero.

REMARK 4.1. Condition 4.3 implies that no 2 reactions in Γ change reactants identically, hence the reactions in Γ are identified uniquely by their consumption of reactants. Condition 4.3 will be satisfied with $\Gamma = \mathcal{M}$ by most SKMs of interest, possibly after explicit inclusion of enzymes in reaction mechanisms. Although autocatalytic reactions such as $X_j + X_k \to^m 2X_k$ and its reverse violate condition (i), these could be accommodated by instead including a more detailed mechanism, for example, $X_j + X_k \to^{m_1} X_j X_k$ and $X_j X_k \to^{m_2} 2X_k$.

An alternative approach would be to work with $G_f^{\sim}$, the graph obtained from the undirected version of the KIG by adding an edge $j \sim k$ whenever $[S_{jm} > 0$ and $S_{km} > 0$ for some $m \in \mathcal{M}]$ and $j \nsim k$ in $G^{\sim}$. The graph $G_f^{\sim}$ might be termed the fraternized (as distinct from moralized) version of the KIG. The separation $A \perp_{G_f^{\sim}} B|D$ implies that $\Delta(A) \cap \Delta(B) = \varnothing$ [since $A \perp_{G^{\sim}} B|D$ and hence $R[m] \subseteq D$ for any $m \in \Delta(A) \cap \Delta(B)$]. Therefore, if $A \perp_{G^{\sim}} B|D$ is replaced by $A \perp_{G_f^{\sim}} B|D$, Condition 4.3 for $\Gamma = \Delta(A) \cap \Delta(B)$ can be dropped from the statements of Theorems 4.4 and 4.5, and from that of Corollary 4.6.

4.2. *Global independence theorems.* We are now in a position to state the main results of Section 4 of the paper. Theorem 4.4 is concerned with global dynamic independence under $\tilde{\mathsf{P}}$, the law of the M-variate Poisson process (see Lemma 4.2).

THEOREM 4.4. *Let G be the KIG of a standard SKM, $[N, S, \mathsf{P}]$, and let $[A, B, D]$ be a partition of $\mathcal{V}$. Suppose also that Condition 4.3 holds for $\Gamma = \Delta(A) \cap \Delta(B)$ (where Γ is possibly empty, in which case the condition is trivial). Then $A \perp_{G^{\sim}} B|D$ implies that $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B | \mathcal{F}_t^{D^*}; \tilde{\mathsf{P}}_t$, where $\{\mathcal{F}_t^{D^*}\}$ is the natural filtration of $N^{D^*}(t)$, $N^{D^*}(t)$ is given by Definition 4.2, and $\tilde{\mathsf{P}}$ is the law of the M-variate Poisson process in Lemma 4.2.*

We provide here a somewhat heuristic discussion of this result, a rigorous treatment being given in Appendix A.1. The argument can be broken down into four steps.

First, the reaction counting processes $[N_m; m = 1, \ldots, M]$ are independent under $\tilde{\mathsf{P}}$. Therefore,

$$\mathcal{F}_t^{\Delta(A)} \perp\!\!\!\perp \mathcal{F}_t^{\Delta(B)\setminus\Delta(A)} | \mathcal{F}_t^{\Delta D_D}; \tilde{\mathsf{P}}_t, \tag{4.2}$$

since $[\Delta(A), \Delta(B) \setminus \Delta(A), \Delta D_D]$ is a partition of the reaction set $\mathcal{M}$. Equation (4.2) holds because the three MVCPs associated with each element of the partition are (unconditionally) independent.

Second, consider again Definition 4.2 for $N^{D^*}(t) = [\{N_A^D(t), N_{AB}^D(t)\}, \{N_B^D(t)\}, \{N_D^D(t)\}]$. The internal history of the first component MVCP in curly parentheses must be contained in the internal history of $N^{\Delta(A)}(t)$. [All the reactions involved in that component change A and hence the sample path of $N^{\Delta(A)}(t)$ implies that of the first component.] Similarly, the internal history of the second component of $N^{D^*}(t)$ in curly parentheses must be contained in that of $N^{\Delta(B)\setminus\Delta(A)}(t)$. The internal history of the third component is equal to $\mathcal{F}_t^{\Delta D_D}$. Combining the internal histories of these 3 components making up $N^{D^*}(t)$ must give $\mathcal{F}_t^{D^*}$. Therefore, the internal histories of the first 2 components can be used to expand the conditioning information in (4.2) to give

$$\mathcal{F}_t^{\Delta(A)} \perp\!\!\!\perp \mathcal{F}_t^{\Delta(B)\setminus\Delta(A)} | \mathcal{F}_t^{D^*}; \tilde{\mathsf{P}}_t. \tag{4.3}$$

Third, establishing the property $\mathcal{F}_t^{\Delta(A)} \perp\!\!\!\perp \mathcal{F}_t^{\Delta(B)} | \mathcal{F}_t^{D^*}; \tilde{\mathsf{P}}_t$ implies the global dynamic independence in Theorem 4.4 since the internal history of the subprocess for A must be contained in that of $N^{\Delta(A)}$ [since the sample path of $N^{\Delta(A)}(t)$ obviously implies that of $N^A(t)$], and similarly for B. This property in turn follows by showing that the internal history of $N^{\Delta(A)\cap\Delta(B)}(t)$ is contained in $\mathcal{F}_t^{D^*}$. The second σ-field in (4.3) can then be expanded to include $\mathcal{F}_t^{\Delta(A)\cap\Delta(B)}$. Combining the internal histories $\mathcal{F}_t^{\Delta(A)\cap\Delta(B)}$ and $\mathcal{F}_t^{\Delta(B)\setminus\Delta(A)}$ in this way gives $\mathcal{F}_t^{\Delta(B)}$.

Finally, $\mathcal{F}_t^{\Delta(A)\cap\Delta(B)} \subseteq \mathcal{F}_t^{D^*}$ is a direct consequence of the fact that $\Delta D_{AB} = \Delta(A) \cap \Delta(B)$—that is, all reactions that change A and B also change D—and that the reactions in ΔD_{AB} change D uniquely (among themselves). These properties of ΔD_{AB} depend crucially on the graphical separation $A \perp_{G^\sim} B | D$ and also on Condition 4.3 holding for ΔD_{AB} (under the conditions of Theorem 4.4). The separation ensures that for any $m \in \Delta(A) \cap \Delta(B)$, the reactants of m are all in D (otherwise, we would have either $A \to B$ or $B \to A$ in the KIG) and hence also $m \in \Delta(D)$. Condition 4.3 ensures that the members of ΔD_{AB} are identified by consumption of reactants, hence the reactions in ΔD_{AB} must change D uniquely (among themselves) and $\mathcal{F}_t^{\Delta(A)\cap\Delta(B)} = \mathcal{F}_{AB}^D(t)$. Therefore, $\mathcal{F}_t^{\Delta(A)\cap\Delta(B)} \subseteq \mathcal{F}_t^{D^*}$, since $N_{AB}^D(t)$ is a component of $N^{D^*}(t)$.

We now turn to consider global dynamic independence under P, the law of the SKM.

THEOREM 4.5. *Let G be the KIG of a standard SKM, $[N, S, \mathsf{P}]$, and let $[A, B, D]$ be a partition of $\mathcal{V}$. Suppose also that Condition 4.3 holds for $\Gamma = \Delta(A) \cap \Delta(B)$ (where Γ is possibly empty, in which case the condition is trivial). Then $A \perp_{G^\sim} B|D$ implies that*

$$\mathcal{L}_t := (\mathsf{dP}_t/\mathsf{d}\tilde{\mathsf{P}}_t)|_{\mathcal{F}_t^N} = \psi_{AD^*,t} \cdot \psi_{BD^*,t}, \qquad t \geq 0,$$

where $\psi_{iD^*,t}$ is a nonnegative, $\mathcal{F}_t^i \vee \mathcal{F}_t^{D^*}$-measurable random variable for $i \in \{A, B\}$, and $\{\mathcal{F}_t^{D^*}\}$ is the natural filtration of $N^{D^*}(t)$.

Taking the logarithm of the likelihood in (4.1) yields

$$\begin{aligned} \log \mathcal{L}_t &= \sum_{m \in \mathcal{M}} \left[t - \int_0^t \lambda_m(u)\, du + \sum_{s \geq 1} 1(T_s^m \leq t) \log(\lambda_m(T_s^m)) \right] \\ &:= \sum_{m \in \mathcal{M}} l_m(t). \end{aligned} \tag{4.4}$$

Theorem 4.5 may be established by showing that, $\forall m \in \mathcal{M}$, $l_m(t)$ is measurable either $\mathcal{F}_t^{AD^*} := \mathcal{F}_t^A \vee \mathcal{F}_t^{D^*}$ or $\mathcal{F}_t^{BD^*} := \mathcal{F}_t^B \vee \mathcal{F}_t^{D^*}$. We explain here how $l_m(t)$ may be computed ($\forall m \in \mathcal{M}$) using either just the sample paths of $N^A(u)$ and $N^{D^*}(u)$, or just the sample paths of $N^B(u)$ and $N^{D^*}(u)$. It is clear from (4.4) that $l_m(t)$ may be computed when $\lambda_m(u)$ may be computed for all $u \in (0, t]$ *and* the sample path of the counting process for that reaction, $N_m(u)$, may be computed over the same time interval (so that the jump times $\{T_s^m \leq t\}$ are known). There are two main elements involved in the argument.

First, the graphical separation $A \perp_{G^\sim} B|D$ again has an important implication for reactants: for any reaction m, either $R[m] \subseteq A \cup D$ or $R[m] \subseteq B \cup D$. Recalling (2.6), only the sample path of the subprocess for the reactants $R[m]$ is needed to compute $\lambda_m(u)$, hence the sample paths of the subprocesses for either $[A, D]$ or $[B, D]$ suffice, according to whether $R[m] \subseteq A \cup D$ or $R[m] \subseteq B \cup D$. [The sample path of D can clearly be computed from that of $N^{D^*}(u)$.]

Second, the sample path $(N_m(u); u \leq t)$ may be computed using just the sample paths of $[N^A, N^{D^*}]$ or $[N^B, N^{D^*}]$, again according to whether $R[m] \subseteq A \cup D$ or $R[m] \subseteq B \cup D$. To see this, consider each group of reactions in the partition of $\mathcal{M}$ given by $[\Delta D_D, \Delta D_{AB}, \Delta(A) \setminus \Delta(B), \Delta(B) \setminus \Delta(A)]$, beginning with $m \in \Delta D_D$. By definition the path of $N^{D^*}(u)$, specifically of its subcomponent $N_D^D(u)$, allows identification of the jump times corresponding to all reactions in ΔD_D that change D identically to m. But since such reactions in ΔD_D change D alone, they must do so uniquely (among reactions in ΔD_D) since no 2 columns of S are equal (Definition 2.1). Therefore, the path of $N^{D^*}(u)$ suffices in this case to compute $(N_m(u); u \leq t)$.

The argument for other groups in the partition is similar. For $m \in \Delta D_{AB}$, it has already been noted that the reactions in ΔD_{AB} change D uniquely (among

themselves). The argument for the last 2 groups is essentially the same. The third group is further partitioned as $[\Delta D_A, \Delta^*(A)]$, where $\Delta^*(A)$ are the reactions that change A alone. Consider $m \in \Delta D_A$—again, by definition, the path of $N^{D^*}(u)$ [specifically, of $N_A^D(u)$] allows identification of the jump times corresponding to the subset of reactions in ΔD_A that change D identically to m. This subset may now contain more than 1 reaction, but inspection of the value of the jumps in the sample path of the subprocess for $[A \cup D]$ corresponding to the jump times so identified allows one to "isolate" just those caused by reaction m (since, again, reactions in ΔD_A change $A \cup D$ uniquely among themselves). The argument for $m \in \Delta^*(A)$ is similar, after noting that the jump times of all reactions in $\Delta^*(A)$ can be identified by eliminating all those of N_A^D and of N_{AB}^D.

The preceding two theorems allow the use of Lemma 4.1 to obtain the following corollary, which summarizes the main results of Section 4.

COROLLARY 4.6. *Let G be the KIG of a standard SKM, $[N, S, \mathsf{P}]$, and let $[A, B, D]$ be a partition of $\mathcal{V}$. Suppose also that Condition 4.3 holds for $\Gamma = \Delta(A) \cap \Delta(B)$ (where Γ is possibly empty). Then the separation $A \perp_{G^\sim} B|D$ in the undirected KIG implies that the global conditional independence $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B|\mathcal{F}_t^{D^*}; \mathsf{P}_t$ holds $\forall t \geq 0$, where $\{\mathcal{F}_t^{D^*}\}$ is the natural filtration of $N^{D^*}(t)$.*

PROOF. Apply Lemma 4.1 to the 3 σ-fields $\mathcal{F}_t^A, \mathcal{F}_t^B, \mathcal{F}_t^{D^*} \subseteq \mathcal{F}_t^N$, recalling from Lemma 4.2 that $\mathsf{P}_t \ll \tilde{\mathsf{P}}_t$. Since $A \perp_{G^\sim} B|D$, Theorem 4.4 implies that $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B|\mathcal{F}_t^{D^*}; \tilde{\mathsf{P}}_t$. Now $\mathcal{F}_t^A \vee \mathcal{F}_t^B \vee \mathcal{F}_t^{D^*} = \mathcal{F}_t^N$, whence $(\mathsf{d}\mathsf{P}_t/\mathsf{d}\tilde{\mathsf{P}}_t)|_{\mathcal{F}_t^A \vee \mathcal{F}_t^B \vee \mathcal{F}_t^{D^*}} = \mathcal{L}_t$, which is given by (4.1). Again since $A \perp_{G^\sim} B|D$, Theorem 4.5 implies that $\mathcal{L}_t = \psi_{AD^*,t} \cdot \psi_{BD^*,t}$, where $\psi_{iD^*,t}$ is a nonnegative, $\mathcal{F}_t^i \vee \mathcal{F}_t^{D^*}$-measurable random variable for $i \in \{A, B\}$. Lemma 4.1 then implies that $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B|\mathcal{F}_t^{D^*}; \mathsf{P}_t$, as required. □

Under the conditions of Corollary 4.6, the separation $A \perp_{G^\sim} B|D$ does *not* imply in general that $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B|\mathcal{F}_t^D; \mathsf{P}_t$, where the conditioning is now on $\mathcal{F}_t^D$ rather than $\mathcal{F}_t^{D^*}$. Similarly, the separation in the moral graph, $A \perp_{G^m} B|D$, does not imply that $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B|\mathcal{F}_t^D; \mathsf{P}_t$. The following theorem and proof establishes both points. The procedure for constructing G^m is the usual one—edges are inserted in the KIG whenever 2 parent nodes of a common child are "unmarried" (i.e., have no edge between them) and then the undirected version of the resulting graph is formed.

THEOREM 4.7. *Let G be the KIG of a standard SKM, $[N, S, \mathsf{P}]$, and let $[A, B, D]$ be a partition of $\mathcal{V}$. Suppose also that Condition 4.3 holds for $\Gamma = \Delta(A) \cap \Delta(B)$ and $A \perp_{G^\sim} B|D$. Then it is possible that neither $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B|\mathcal{F}_t^D; \tilde{\mathsf{P}}_t$ nor $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B|\mathcal{F}_t^D; \mathsf{P}_t$ holds, where $\{\mathcal{F}_t^D\}$ is as usual the internal history of the subprocess $N^D(t)$.*

PROOF. The proof is by example. Consider the standard SKM with $\mathcal{V} = \{A, B, D\}$ and reactions

$$A \to^{f} D, \qquad D \to^{r} A, \qquad D \to^{irr} B,$$

which has the KIG, $G = A \longleftrightarrow D \to B$. Note that $G^{\sim} = G^{m}$. Clearly, $\Gamma = \varnothing$ and $A \perp_{G^{\sim}} B|D$. Note also that $N^{A}(t) = [N_f(t), N_r(t)]'$, $N^{D}(t) = [N_f(t), N_r(t) + N_{irr}(t)]'$ and $X^{B}(t) - X^{B}(0) = N_{irr}(t)$. It suffices to show that, under both $\tilde{\mathsf{P}}_t$ and P_t, $\mathsf{E}[X^{B}(t) - X^{B}(0)|\mathcal{F}_t^{D}]$ is not a version of $\mathsf{E}[X^{B}(t) - X^{B}(0)|\mathcal{F}_t^{A} \vee \mathcal{F}_t^{D}]$. First, show that $\mathcal{F}_t^{A} \vee \mathcal{F}_t^{D} = \mathcal{F}_t^{N}$. Clearly, $\mathcal{F}_t^{A} \vee \mathcal{F}_t^{D} \subseteq \mathcal{F}_t^{N}$, and since $N_{irr}(s) = [N_r(s) + N_{irr}(s)] - N_r(s)$, $N_{irr}(s)$ is measurable $\mathcal{F}_t^{A} \vee \mathcal{F}_t^{D}$, hence $\mathcal{F}_t^{N} \subseteq \mathcal{F}_t^{A} \vee \mathcal{F}_t^{D}$. It follows that, under both $\tilde{\mathsf{P}}_t$ and P_t, $\mathsf{E}[X^{B}(t) - X^{B}(0)|\mathcal{F}_t^{A} \vee \mathcal{F}_t^{D}] = N_{irr}(t)$ since $N_{irr}(t)$ is measurable $\mathcal{F}_t^{A} \vee \mathcal{F}_t^{D}$. However, $N_{irr}(t)$ is clearly not measurable $\mathcal{F}_t^{D}$ and so cannot be a version of $\mathsf{E}[X^{B}(t) - X^{B}(0)|\mathcal{F}_t^{D}]$ under either probability measure. In fact, it is possible to show that under $\tilde{\mathsf{P}}_t$, $\mathsf{E}[X^{B}(t) - X^{B}(0)|\mathcal{F}_t^{D}] = \frac{1}{2}[N_r(t) + N_{irr}(t)]$. □

4.3. *Histories of the separator, D.* It is of interest in applications to understand, for a given partition $[A, B, D]$ of $\mathcal{V}$, how the histories $\{\mathcal{F}_t^{D^*}\}$ and $\{\mathcal{F}_t^{D}\}$ differ. A comparison of $N^{D^*}(t)$ and $N^{D}(t)$ is equivalent to a comparison of the corresponding partitions of the reactions $\Delta(D)$.

PROPOSITION 4.8. *The partition given by* $\mathcal{M}^*(\Delta(D)) := \{\mathcal{M}(\Delta D_A) \cup \mathcal{M}(\Delta D_{AB}) \cup \mathcal{M}(\Delta D_B) \cup \mathcal{M}(\Delta D_D)\}$ *is a refinement of the partition* $\mathcal{M}(\Delta(D))$, *so that every element of* $\mathcal{M}(\Delta(D))$ *is a union of elements of* $\mathcal{M}^*(\Delta(D))$ (*see Definition 2.2 for the partition notation used*). *Hence,* $\mathcal{F}_t^{D^*} \supseteq \mathcal{F}_t^{D}$ *and* $\mathcal{F}_t^{A} \vee \mathcal{F}_t^{B} \vee \mathcal{F}_t^{D^*} = \mathcal{F}_t^{N}$ $\forall t$.

PROOF. Take an element of $\mathcal{M}(\Delta(D))$, $\mathcal{M}_e(\Delta(D))$ say. Let $m \in \mathcal{M}_e(\Delta(D))$ and denote the element of $\mathcal{M}^*(\Delta(D))$ to which m belongs as $\mathcal{M}_m^*(\Delta(D))$. Now $\mathcal{M}_m^*(\Delta(D)) \subseteq \mathcal{M}_e(\Delta(D))$ since all elements of $\mathcal{M}_m^*(\Delta(D))$ change D equivalently (resulting in the same change to D as m does). Thus, $\mathcal{M}_e(\Delta(D)) = \bigcup_{m \in \mathcal{M}_e(\Delta(D))} \mathcal{M}_m^*(\Delta(D))$, which establishes the first claim. It then follows from Definition 2.2 that $\mathcal{F}_t^{D^*} \supseteq \mathcal{F}_t^{D}$ since elements of $N^{D}(t)$ are obtained by summing (where necessary) the appropriate elements of $N^{D^*}(t)$. Lemma 2.1 established that $\mathcal{F}_t^{A} \vee \mathcal{F}_t^{B} \vee \mathcal{F}_t^{D} = \mathcal{F}_t^{N}$. But $\mathcal{F}_t^{D} \subseteq \mathcal{F}_t^{D^*}$ then implies $\mathcal{F}_t^{N} \subseteq \mathcal{F}_t^{A} \vee \mathcal{F}_t^{B} \vee \mathcal{F}_t^{D^*} \subseteq \mathcal{F}_t^{N}$. □

In computational work with SKMs, establishing if the partitions $\mathcal{M}^*(\Delta(D))$ and $\mathcal{M}(\Delta(D))$ are identical provides a straightforward means of checking whether the processes $N^{D^*}(t)$ and $N^{D}(t)$ are identical. The two partitions are identical if and only if there do not exist two reactions in different elements of $[\Delta D_A, \Delta D_{AB}, \Delta D_B, \Delta D_D]$ that result in the same change in D—that is, there

do not exist 2 reactions in $\Delta(D)$ that change D identically but do not have the same membership of both of the sets $[\Delta(A), \Delta(B)]$.

PROPOSITION 4.9. *Let* $[A, B, D]$ *be a partition of* $\mathcal{V}$, *the species set of an SKM. Then* $N^{D^*}(t) = N^D(t)\ \forall t, \forall \omega \in \Omega$, *if and only if the following condition holds*: *for any* 2 *reactions* $m, \tilde{m} \in \Delta(D)$ *with* $S_m^D = S_{\tilde{m}}^D$, *the reaction* m *has the same membership of the two sets* $[\Delta(A), \Delta(B)]$ *as does the reaction* $\tilde{m}$.

Under this condition, $\{\mathcal{F}_t^{D^*}\} = \{\mathcal{F}_t^D\}$.

PROOF. If the condition holds both m and $\tilde{m}$ are members of an equivalence class of some $\Delta D_\bullet$ Hence, any 2 members of an equivalence class of $\Delta(D)$—that is, of an element of $\mathcal{M}(\Delta(D))$—are also both members of an element of $\mathcal{M}^*(\Delta(D))$. Therefore, by Proposition 4.8, $\mathcal{M}(\Delta(D)) = \mathcal{M}^*(\Delta(D))$, whence $N^{D^*}(t) = N^D(t)\ \forall t, \forall \omega$.

Conversely, suppose $N^{D^*}(t) = N^D(t)\ \forall t, \forall \omega$. Then the vectors $N^{D^*}(t) = N^D(t)$ have the same dimension and so $\mathcal{M}^*(\Delta(D))$ cannot be a strict refinement of $\mathcal{M}(\Delta(D))$. Hence, by Proposition 4.8, $\mathcal{M}(\Delta(D)) = \mathcal{M}^*(\Delta(D))$. Suppose the reactions $(m, \tilde{m})$ differ in their membership of the two sets $\Delta(A), \Delta(B)$. Then $(m, \tilde{m})$ are in different elements of $\mathcal{M}^*(\Delta(D))$ but the same element of $\mathcal{M}(\Delta(D))$, which is a contradiction. □

In applications where it is more convenient or practical to include only internal histories, checking the condition of Proposition 4.9—or equivalently, the equality of $\mathcal{M}(\Delta(D))$ and $\mathcal{M}^*(\Delta(D))$—often reveals that the processes $N^{D^*}(t)$ and $N^D(t)$ are similar or identical. This is in part because, in practice, many elements of $\mathcal{M}(\Delta(D))$ are single reactions—that is, many of the reactions that change D are uniquely identified by the corresponding change in D. Furthermore, where $\mathcal{F}_t^D \subset \mathcal{F}_t^{D^*}$ (strictly), the partition $[A, B, D]$ can often be altered slightly to make the processes $N^{D^*}(t)$ and $N^D(t)$ identical. Examples of this are given in Section 6 in connection with the red blood cell SKM.

5. Independence and modularity. Rigorous mathematical definition and identification of modularizations for biochemical reaction networks is recognized as being a difficult problem, especially from a dynamic perspective [29]. A prominent approach has been to construct a graph representing "interactions" between species and to consider different *partitions* of the species between modules, maximizing an objective function based on the fraction of edges that are intra-modular relative to the expected fraction in an "equivalent," randomized graph when the same partition of species is used [13, 18]. From a stochastic process perspective, the graphs used often do not encode properly the dependence structure of the molecular network—for example, in contrast to a KIG, metabolic network graphs typically omit the local dependence between reactants in the same reaction, only capturing that between reactant and product. The approach is intended to operationalize the concept that modules function "near-independently." However, the measure

of modularity adopted for the objective function is rather distant from well-defined notions of dynamic (in)dependence between species. The local and global conditional independence results developed in Sections 3 and 4 make it possible to add content to and make rigorous what is meant by near-independence of modules, and to accommodate "overlapping" modules with nonempty intersection.

The term modularization is derived from the biological literature where "modularity" has been much discussed. A modularization here is a hypergraph of the vertex set of the KIG (i.e., a collection of subsets of species) with the following property—the internal history at time t of each subset (or module) is conditionally independent of the internal history of all the other modules, given the history of its intersection with those modules.

DEFINITION 5.1. Let $\mathcal{V}$ be the species set of an SKM $[N, S, \mathsf{P}]$. The finite collection of subsets of $\mathcal{V}$, $\{M_d | M_d \subseteq \mathcal{V}\}$, is a *modularization of the SKM* if and only if $\bigcup_d M_d = \mathcal{V}$ and

$$\mathcal{F}_t^{M_d} \perp\!\!\!\perp \mathcal{F}_t^{\bigcup_{e\neq d} M_e} | \mathcal{F}_t^{S_d^*}; \mathsf{P} \qquad \forall d, t, \tag{5.1}$$

where $S_d = M_d \cap \{\bigcup_{e\neq d} M_e\}$ and the history $\{\mathcal{F}_t^{S_d^*}\}$ is the natural filtration of $N^{S_d^*}(t)$. The latter is given as usual by Definition 4.2, applied to the partition $[M_d \setminus S_d, \mathcal{V} \setminus M_d, S_d]$.

Note that since $\mathcal{V} \setminus M_d = \{\bigcup_{e\neq d} M_e\} \setminus S_d$, (5.1) is equivalent to the statement $\mathcal{F}_t^{M_d \setminus S_d} \perp\!\!\!\perp \mathcal{F}_t^{\mathcal{V}\setminus M_d} | \mathcal{F}_t^{S_d^*}; \mathsf{P}\ \forall d, t$. Roughly speaking, the global evolution on $[0, t]$ of the species in $M_d \setminus S_d$ and the species in $\mathcal{V} \setminus M_d$ ("the rest of the network") are conditionally independent given the history of the intersection, $\mathcal{F}_t^{S_d^*}$. We will say that two modularizations are *nested* if each module of one of the modularizations is contained in some module of the other modularization.

Of course some modularizations of an SKM will be more useful than others. It will usually be desirable for the intersections S_d to contain a relatively small number of species and to be able to move between nested modularizations, thus considering finer and coarser levels of resolution. Computationally efficient methods are developed below for the identification of such modularizations that are based around the maximal prime decomposition of the undirected version of the KIG of the SKM, $G^{\sim}$. It will be proved below that applying such graphical decomposition methods results in subgraphs whose vertex sets, $\{M_d\}$ say, satisfy the graphical separation $M_d \perp_{G^{\sim}} \bigcup_{e\neq d} M_e | S_d\ \forall d$. Therefore, under the conditions of Corollary 4.6, the required global dynamic independence of (5.1) holds for all d, and $\{M_d\}$ constitutes a modularization according to Definition 5.1.

5.1. *Identifying modularizations by graph decomposition.* Some definitions from the graphical literature will prove useful (for further details, see [21]). An

undirected graph is said to be *complete* if there is an edge between all pairs of vertices in its vertex set. Let H be an undirected graph with vertex set $\mathcal{V}$. The subgraph induced by $M_d \subset \mathcal{V}$, $H(M_d)$, consists of the vertices in M_d and exactly the edges between those vertices that occur in H itself. A partition $[A, B, D]$ of $\mathcal{V}$, $A, B \neq \varnothing$, forms a *decomposition* of H into the subgraphs $H(A \cup D)$ and $H(B \cup D)$ if the separation $A \perp_H B|D$ holds and the subgraph $H(D)$ is complete. The subgraph $H(M_d)$ is *prime* if there does not exist a decomposition of $H(M_d)$.

DEFINITION 5.2. Let H be an undirected graph with vertex set $\mathcal{V}$, and $M_d \subseteq \mathcal{V}$. The induced subgraph $H(M_d)$ is a maximal prime subgraph of H if $H(M_d)$ is prime and there exists a decomposition of $H(N)$ for all N satisfying $M_d \subset N \subseteq \mathcal{V}$. The *maximal prime subgraph decomposition* (*MPD*) of H is given by $\{H(M_d)\}$, the unique collection of maximal prime subgraphs of H, and satisfies that $\bigcup_d M_d = \mathcal{V}$.

A *junction tree* representation of the MPD, $\mathcal{T}_{\text{MPD}}$, always exists and has the subsets $\{M_d\}$ as its clusters (i.e., as the vertices of the junction tree) [24]. A junction tree $\mathcal{T}$ is a connected, undirected graph without cycles in which the intersection of any 2 clusters of the tree, $M_d \cap M_e$ $(d \neq e)$, is contained in every cluster on the unique path in $\mathcal{T}$ between M_d and M_e. Such trees will prove very useful in visualizing, representing and manipulating modularizations of SKMs. We say, for reasons that will become apparent, that any 2 clusters adjacent in the tree are separated by their intersection, and call that intersection a *separator of* $\mathcal{T}$.

The SKM modularization algorithm presented below contains as a special case the method due to [24] for computation of $\mathcal{T}_{\text{MPD}}$, applied to the undirected version of the KIG, $G^{\sim}$. The advantage of this version of Algorithm 5.1 is that it can be fully automated to identify the MPD modularization of the SKM in a manner that is computationally feasible even for very large SKMs. However, it will often be informative to consider a range of nested modularizations in order to explore the different levels of organization of the reaction network. To this end, the general version of Algorithm 5.1 first obtains a junction tree of the clique decomposition for $G_T^{\sim}$ (a minimal triangulation of $G^{\sim}$)—this provides the finest, most detailed modularization that is identified. The clique decomposition of $G_T^{\sim}$ is unique (since it corresponds to the MPD of $G_T^{\sim}$). Coarser-grained modularizations, including the MPD one, are obtained by successively aggregating adjacent clusters in the junction tree.

ALGORITHM 5.1. Let G be the KIG of an SKM.

1. Construct $G^{\sim}$, the undirected version of G;
2. Construct $G_T^{\sim}$, a minimal triangulation of $G^{\sim}$;
3. Obtain the clique decomposition of $G_T^{\sim}$ with the cliques, $\{C_1, C_2, \ldots, C_\delta\}$ say, ordered to satisfy the running intersection property (i.e., for $e = 2, \ldots, \delta$, $\exists d^* \in \{1, \ldots, e-1\}$ s.t. $C_e \cap \{\bigcup_{i=1}^{e-1} C_i\} \subseteq C_{d^*}$);

4. Organize the clique decomposition as a (rooted) junction tree $\mathcal{T}_C$ in which, for $e = 2, \ldots, \delta$, the parent of C_e is C_{d^*}; set $\mathcal{T} = \mathcal{T}_C$;
5. Either go to step 7 or, select a pair of adjacent clusters (C_i, C_j) in $\mathcal{T}$ $(i < j)$ and *aggregate* them by updating $\mathcal{T}$ as follows: set $P = \text{pa}(C_i)$ and $C = \{\text{ch}(C_i) \cup \text{ch}(C_j)\} \setminus C_j$, replace cluster i by $C_i \cup C_j$ (retaining its numbering, i), set $\text{pa}(C_i) = P$, and set $\text{ch}(C_i) = C$;
6. Go to step 5;
7. Return $\mathcal{T}_{\text{MOD}} = \mathcal{T}$.

The property that $G_T^{\sim}$ is triangulated is equivalent to saying that $G_T^{\sim}$ can be decomposed recursively until all the resulting subgraphs are complete [21]. Such a recursive decomposition produces a collection of subgraphs containing the cliques $\{G_T^{\sim}(C_d)\}$, that is, the maximally complete subgraphs of $G_T^{\sim}$. Triangulation refers to the operation of adding edges to $G^{\sim}$ so that it becomes triangulated. The triangulation $G_T^{\sim}$ in step 2 must be minimal—that is, one for which removal of any edge added during triangulation results in an untriangulated graph—otherwise, Remark 5.1 below does not hold, in general.

Efficient algorithms have been developed in the graphical literature for both minimal triangulation and clique decomposition (see [3, 24]) which can be exploited here to compute the SKM modularizations and associated junction trees. The following special case of Algorithm 5.1 returns the junction tree representation of the maximal prime decomposition (MPD) of the undirected KIG, $G^{\sim}$ [24].

REMARK 5.1. Algorithm 5.1 returns $\mathcal{T}_{\text{MPD}}$ for the undirected KIG, $G^{\sim}$, when step 5 is replaced by:

5′. While [there exists a separator S of $\mathcal{T}$ such that $G^{\sim}(S)$ is incomplete], aggregate within $\mathcal{T}$ the 2 clusters separated by S; then go to step 7.

It is worth noting the time complexity of steps 2 and 4. The general problem of finding an optimal triangulation of an undirected graph (i.e., one that adds least edges among all triangulations) is *NP-hard*. The complexity of minimal triangulation (step 2) is $\mathcal{O}(ne)$ where e is the number of edges in $G^{\sim}$, [24]. The complexity of constructing the clique junction tree $\mathcal{T}_C$ (steps 3 and 4 combined) is $\mathcal{O}(n^2)$, [24].

5.2. *Nested modularizations and junction trees.* A concise proof that the clusters of the tree $\mathcal{T}_{\text{MOD}}$ returned by Algorithm 5.1 constitute a modularization of the SKM—with any choice of aggregation scheme in stage 5—is made possible by establishing that $\mathcal{T}_{\text{MOD}}$, like $\mathcal{T}_C$, is a junction tree, and that the intersections of adjacent clusters of $\mathcal{T}_{\text{MOD}}$ continue to correspond to separators in $G_T^{\sim}$, and hence in $G^{\sim}$. The following proposition does just that.

PROPOSITION 5.2. *Let $\mathcal{T}_{\text{MOD}}$ be the undirected graph returned by applying Algorithm 5.1 to the KIG, G, of an SKM. Denote the clusters (modules) of $\mathcal{T}_{\text{MOD}}$*

by $\{M_d\}$. *Then* $\mathcal{T}_{\mathrm{MOD}}$ *is a junction tree. Suppose that* (M_d, M_e) *are any* 2 *adjacent clusters in* $\mathcal{T}_{\mathrm{MOD}}$ *with separator* $S_{de} := M_d \cap M_e$, *and that* (*as is conventional*) *the edges* $M_d \sim M_e$ *are labeled by the corresponding separator* S_{de}.

Then $S_{de} = V_{de} \cap V_{ed}$ *and the graphical separation* $V_{de} \perp V_{ed} | S_{de}$ *holds in* $G_T^{\sim}$, *and hence in* $G^{\sim}$, *where* V_{de} (V_{ed}) *is the union of the clusters in* $\mathcal{T}^{de}_{\mathrm{MOD}}$ ($\mathcal{T}^{ed}_{\mathrm{MOD}}$), *the* $\mathcal{T}^{\bullet}_{\mathrm{MOD}}$ *are the* 2 *subtrees obtained by cutting the edge* $M_d \sim M_e$ *in* $\mathcal{T}_{\mathrm{MOD}}$, *and* $M_d \subseteq V_{de}$ ($M_e \subseteq V_{ed}$).

Proof of Proposition 5.2 is given in Appendix B.

We can now state and prove the result that establishes the validity of our modularization identification methods.

THEOREM 5.3. *Let* G *be the KIG of a standard SKM,* $[N, S, \mathsf{P}]$, *and let* $\mathcal{T}_{\mathrm{MOD}}$ *be the junction tree of modules,* $\{M_d\}$, *returned by Algorithm* 5.1. *Suppose also that Condition* 4.3 *holds for* $\Gamma_d = \Delta(M_d \setminus S_d) \cap \Delta(\mathcal{V} \setminus M_d)$ $\forall d$. *Then* $\{M_d\}$ *is a modularization of the SKM in the sense of Definition* 5.1; *and each* S_d *is given by* $\bigcup_{e \in ne(M_d)} S_{de}$ [*where* $ne(M_d)$ *is the indices of those clusters that have edges with* M_d *in* $\mathcal{T}_{\mathrm{MOD}}$]. *Furthermore, each module residual* $M_d \setminus S_d$ *is locally independent of* $\mathcal{V} \setminus M_d$ *given the internal history of* M_d.

PROOF. By Corollary 4.6, it suffices to show that the separation $\{M_d \setminus S_d\} \perp_{G_T^{\sim}} \{\mathcal{V} \setminus M_d\} | S_d$ holds in $G_T^{\sim}$, for all d, since then $\{M_d \setminus S_d\} \perp_{G^{\sim}} \{\mathcal{V} \setminus M_d\} | S_d$ holds in the undirected KIG $G^{\sim}$. This follows because every path in $G^{\sim}$ from $M_d \setminus S_d$ to $\mathcal{V} \setminus M_d$ is also such a path in $G_T^{\sim}$. Recall that by definition $S_{de} := M_d \cap M_e$. Hence,

$$
\begin{aligned}
S_d &= \Big\{ \bigcup_{e \in ne(M_d)} (M_d \cap M_e) \Big\} \cup \Big\{ \bigcup_{e \notin ne(M_d)} (M_d \cap M_e) \Big\} \\
&= \bigcup_{e \in ne(M_d)} (M_d \cap M_e) = \bigcup_{e \in ne(M_d)} S_{de},
\end{aligned} \tag{5.2}
$$

where the second line holds by the fact that $\mathcal{T}_{\mathrm{MOD}}$ is a junction tree (Proposition 5.2) since, for $e \notin ne(M_d)$, $(M_d \cap M_e)$ is contained in $M_{\tilde{e}}$, and thus in $S_{d\tilde{e}}$ for some $\tilde{e} \in ne(M_d)$ lying on the unique path between M_d and M_e in $\mathcal{T}_{\mathrm{MOD}}$.

By Proposition 5.2, $M_d \perp_{G_T^{\sim}} V_{ed} | S_{de}$ $\forall e \in ne(M_d)$, since $M_d \subseteq V_{de}$. Hence, $M_d \perp_{G_T^{\sim}} V_{ed} | \{\bigcup_{e \in ne(M_d)} S_{de}\}$ and, since this holds for all $e \in ne(M_d)$, $M_d \perp_{G_T^{\sim}} \{\bigcup_{e \in ne(M_d)} V_{ed}\} | \{\bigcup_{e \in ne(M_d)} S_{de}\}$. Now $\{\bigcup_{e \in ne(M_d)} V_{ed}\} = \{\bigcup_{e \neq d} M_e\}$. To see this, note that the latter is the union of those clusters reachable by paths in $\mathcal{T}_{\mathrm{MOD}}$ that start with the edge $M_d \sim M_e$ for some $e \in ne(M_d)$ (since $\mathcal{T}_{\mathrm{MOD}}$ is connected); and V_{ed} is the union of those clusters reachable by paths in $\mathcal{T}_{\mathrm{MOD}}$ that start at the node M_e (since $\mathcal{T}^{ed}_{\mathrm{MOD}}$ is connected). Therefore, using (5.2), $M_d \perp_{G_T^{\sim}} \{\bigcup_{e \neq d} M_e\} | S_d$, as required. □

6. SKM of red blood cell. We now apply the modularization techniques of Section 5 and the underlying dynamic independence theory on which they are based to identify biologically interesting modulariszations of an SKM of the human red blood cell. The study of this metabolic reaction network was an early success of a systems biology approach [19, 27]. There now exists detailed knowledge of the component reactions as a result of at least three decades of research on both the biochemical and mathematical modeling fronts. The identification of aggregates of metabolites (i.e., species) and regulatory structures in the red blood cell has also received attention from a systems biology perspective [19, 25]. This particular reaction network therefore constitutes a suitable test-bed to establish the utility and applicability of our approach. In contrast to this work, [19] aims to identify "pools" of metabolites in the red blood cell, that is "aggregate groups of [species] which [...] move together in a concerted manner," rather than groups that move independently given an appropriate conditioning set of species.

The SKM studied is the one implied by the metabolic network of the red blood cell published in the open access Biomodels Database [23], which in turn is a slightly extended version of the kinetic model of [27] and [14]. The SKM consists of 38 reactions, with 45 different biochemical species in the species set $\mathcal{V}$ (the enzymes, i.e., catalysts, involved are omitted from $\mathcal{V}$ as they do not appear explicitly in the reaction mechanisms). Full details are available from [23]. The direction of the reactions is as for the kinetic model in Table 1 of [14], except for 8 additional reactions which are all included as dissociation reactions. It was verified that the SKM, henceforth $\mathcal{SKM}_{\text{rbc}}$, is a standard SKM (according to Definition 4.3). The names of the biochemical species in $\mathcal{V}$ and the associated abbreviations used are given in Appendix C. For details of the reactions in $\mathcal{M}$, the reader is referred to [23].

Figure 2 depicts the kinetic independence graph G for $\mathcal{SKM}_{\text{rbc}}$. The graph is a powerful visual aid to understanding the architecture of the molecular network and can be preliminarily inspected for interesting local independences and separations in the undirected version $G^{\sim}$. The clique decomposition, $\mathcal{T}_C$, from Algorithm 5.1 for $\mathcal{SKM}_{\text{rbc}}$ has many clusters (20 out of 38) for which $M_d \setminus S_d$ is the empty set. It is therefore desirable to implement Algorithm 5.1 with a substantial degree of pairwise cluster aggregation in step 5. On the other hand, $\mathcal{T}_{\text{MPD}}$ for this SKM is overly coarse-grained for most purposes. Figure 3 depicts a particular junction tree $\mathcal{T}_{\text{MOD},1}$ returned by Algorithm 5.1, with the choice of aggregations guided both by the structure of $\mathcal{T}_C$ itself and the goal of a modularisation that offers biological insight. This approach relies on and takes advantage of the flexibility offered by Algorithm 5.1—some exploration of alternative modularizations by the user is required, but no prior information about possible modularizations is needed.

The junction tree $\mathcal{T}_{\text{MOD},1}$ in Figure 3 is labelled as follows. The dth module (rectangle) is labeled with the species in the "residual module," $M_d \setminus S_d$, and each edge, $M_d \sim M_e$, is labeled with the species in the separator S_{de}, that is the intersection of the modules connected by that edge. It was verified that, for all d,

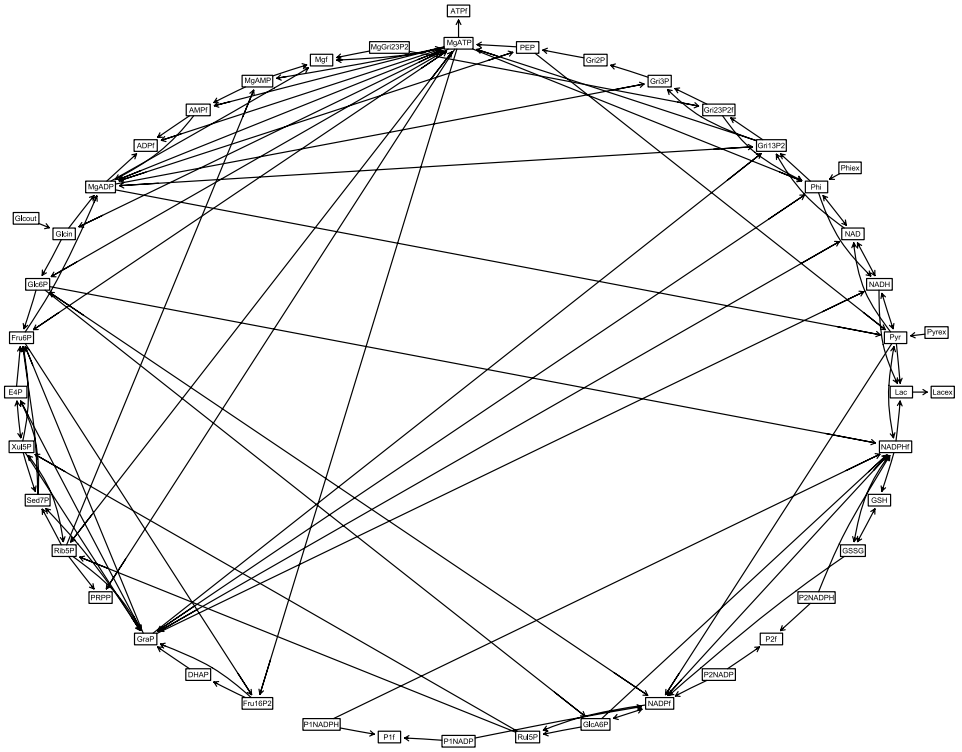

FIG. 2. *Kinetic independence graph (KIG) of* $\mathcal{SKM}_{\mathrm{rbc}}$, *the Metabolic Network of the Human Red Blood Cell* [23]. *The KIG is constructed according to Definition* 3.1. *Full species names are given in Appendix* C.

Condition 4.3 holds for $\Gamma_d = \Delta(M_d \setminus S_d) \cap \Delta(\mathcal{V} \setminus M_d)$, as required by Theorem 5.3. Recall that Proposition 5.2 implies $S_{de} = V_{de} \cap V_{ed}$ and $V_{de} \perp_{G^\sim} V_{ed} | S_{de}$ in $G^\sim$. Such separations may be conveniently read off from any junction tree $\mathcal{T}_{\mathrm{MOD}}$ since S_{de}, V_{de} and V_{ed} are all immediately apparent from examination of the tree. Similarly, the defining conditional independencies of the modularization, namely $\mathcal{F}_t^{M_d \setminus S_d} \perp\!\!\!\perp \mathcal{F}_t^{\{\bigcup_{e \neq d} M_e\} \setminus S_d} | \mathcal{F}_t^{S_d^*}; \mathsf{P}\ \forall d$ (5.1), may be read off the junction tree using Theorem 5.3, S_d being given by the union of the labels of all edges that connect with the dth module.

Having obtained a modularization such as $\mathcal{T}_{\mathrm{MOD},1}$, the next stage is to ask what are the interesting features that emerge from a biochemical and systems biological perspective. Each of the main modules of $\mathcal{T}_{\mathrm{MOD},1}$ turns out to contain like species, either in terms of their molecular structure (e.g., the groupings of monosaccharide-phosphate sugar molecules and phosphoglycerate molecules) or their function (e.g., the grouping of species involved in reduction–oxidation reactions), or both.

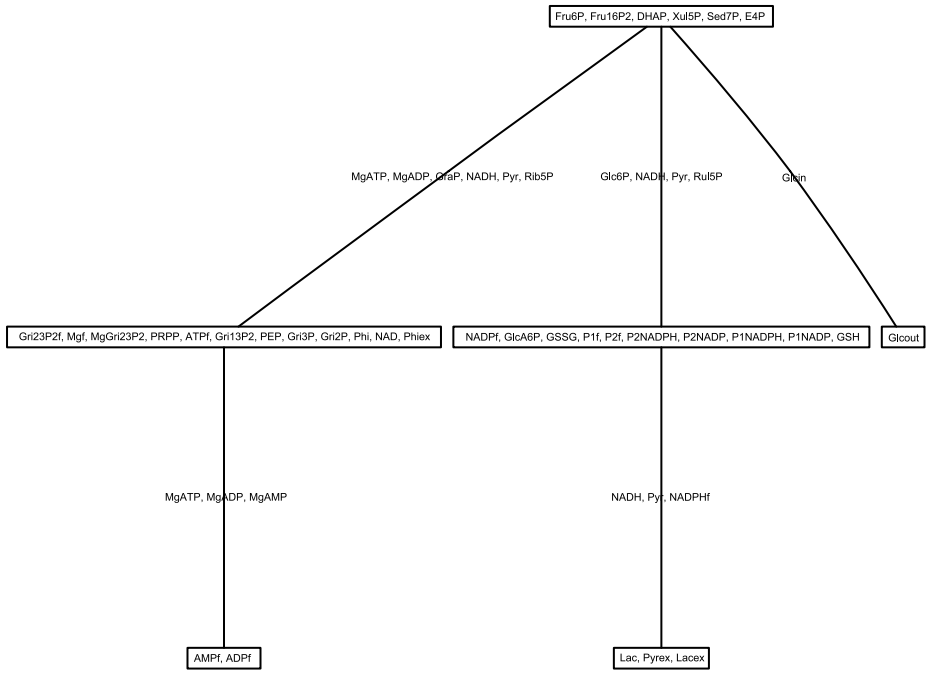

FIG. 3. *Junction tree representation*, $\mathcal{T}_{\mathrm{MOD},1}$, *of a modularisation of* $\mathcal{SKM}_{\mathrm{rbc}}$, *the Metabolic Network of the Human Red Blood Cell* [23]. *The global dynamic independences* $\mathcal{F}_t^{M_d} \perp\!\!\!\perp \mathcal{F}_t^{\bigcup_{e\neq d} M_e} | \mathcal{F}_t^{S_d^*}$ *hold for each module* M_d (*see Definition* 5.1). *The modules* (*rectangles*) *are labeled with their residuals and edges are labeled with the intersection of adjacent modules. Full species names are given in Appendix* C.

Specific modules and their residuals are denoted by their first constituent species in the subsequent discussion. Consider first the central residual $\{NADPf, \ldots\}$ in $\mathcal{T}_{\mathrm{MOD},1}$, the residual of what will be termed the *Redox module* (for *Red*uction-*ox*idation). The red blood cell is subject to oxidative stress due to reactive oxygen species, which if left unchecked leads to cell lysis (bursting) and consequent anemia. All of this residual's species can be seen to play a role in the control of such oxidative stress. Glutathione (*GSH*) acts as an antioxidant, scavenging reactive oxygen species and itself being oxidised as a result (giving rise to the reaction $2GSH \rightarrow GSSG$). The cell must maintain adequate levels of *GSH*, which it does by producing large amounts of *NADPH* for use in the reduction of *GSSG* (by the reaction $GSSG + NADPH \rightarrow 2GSH + NADP$). Production of *NADPH* is via 2 reactions (usually described as the oxidative phase of the pentose phosphate pathway), both of which involve *GlcA6P*. Both *NADP* and *NADPH* are also found bound to the proteins $P1$ and $P2$. Notice that the reduced forms *NADPH* and *NADH* are both found in the module's separator (edge) with $\{Lac, Pyrex, Lacex\}$, since both influence the intensity of lactate (*Lac*) production and export as reactants for the reduction of pyruvate (*Pyr*).

The module $\{NADPf, \ldots\}$ clearly has an important function in oxidative stress control and in reduction–oxidation reactions more generally within the red blood cell. Of course, these functions of its individual species are well known. That their dynamic evolution, together with that of lactate (*Lac*), is globally independent of all the other species in the network conditional on the internal history of {*Fru6P*, *Glc6P*, *NADH*, *Pyr*, *Rul5P*, *Pyrex*} is an insight provided by the modularizations (see also the derivation of $\mathcal{T}_{\mathrm{MOD},2}$ below). Assigning function(s) where possible to each module of a given modularization, $\mathcal{T}_{\mathrm{MOD}}$, is likely to improve both understanding of a network and ultimately aid attempts to control it. For reasons of space, comments related to the remaining two large residuals of $\mathcal{T}_{\mathrm{MOD},1}$ may be found as part of the discussion of $\mathcal{T}_{\mathrm{MOD},2}$ below.

The structure of $\mathcal{T}_{\mathrm{MOD},1}$ encourages further aggregation in an obvious manner. A second modularization, $\mathcal{T}_{\mathrm{MOD},2}$, of $\mathcal{SKM}_{\mathrm{rbc}}$ is thus shown in Figure 4. [It was verified that, in this case also, Condition 4.3 holds for $\Gamma_d = \Delta(M_d \setminus S_d) \cap \Delta(\mathcal{V} \setminus M_d)\ \forall d$.] $\mathcal{T}_{\mathrm{MOD},2}$ may be derived from $\mathcal{T}_{\mathrm{MOD},1}$in two steps. First, the modules $\{AMPf, ADPf\}$, $\{Lac, Pyrex, Lacex\}$ and $\{Glcout\}$ are aggregated with their adjacent modules in $\mathcal{T}_{\mathrm{MOD},1}$. Second, a small number of species in residuals are then

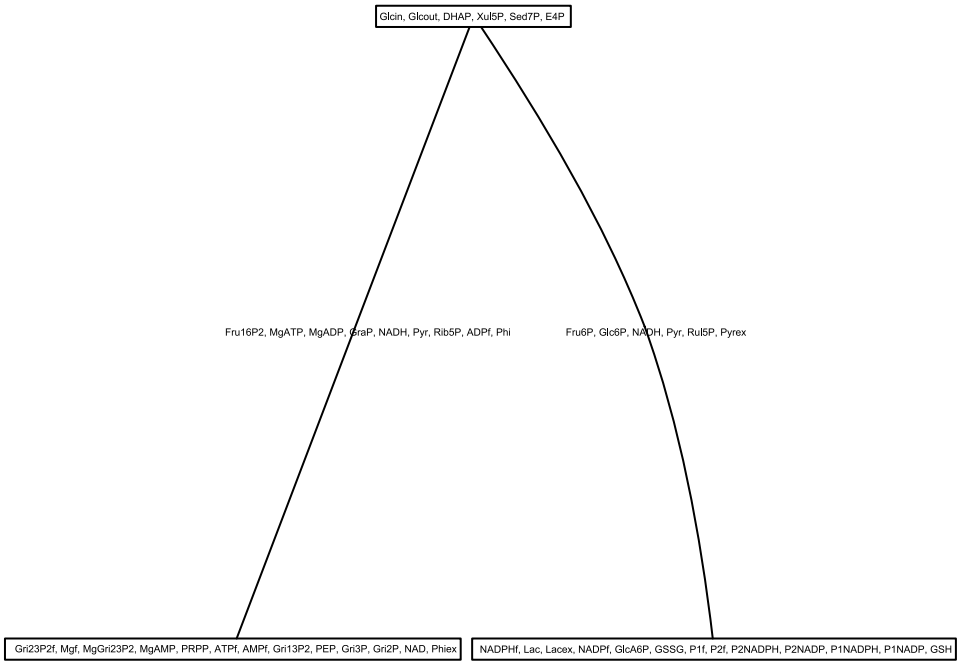

FIG. 4. *Junction tree representation, $\mathcal{T}_{\mathrm{MOD},2}$, a coarser-grained modularization of $\mathcal{SKM}_{\mathrm{rbc}}$, the Metabolic Network of the Human Red Blood Cell* [23]. *The global dynamic independences $\mathcal{F}_t^{M_d} \perp\!\!\!\perp \mathcal{F}_t^{\bigcup_{e\neq d} M_e} | \mathcal{F}_t^{S_d}$ hold for each module M_d (see Definition 5.1). The rectangles contain module residuals and edges are labeled with the intersection of adjacent modules. Full species names are given in Appendix C.*

judiciously included also in an additional module so that they both fall instead in the relevant separator and the condition of Proposition 4.9 is satisfied for each partition $[M_d \setminus S_d, \mathcal{V} \setminus M_d, S_d]$. By Proposition 4.9, this ensures that N^{S_d} and $N^{S_d^*}$ are the same subprocess, whence $\{\mathcal{F}_t^{S_d}\} = \{\mathcal{F}_t^{S_d^*}\}$ for $d = 1, 2, 3$. Clearly, the second step need only be performed if it is desired to be able to replace $\mathcal{F}_t^{S_d^*}$ by $\mathcal{F}_t^{S_d}$ in the defining conditional independencies of the modularization (Definition 5.1). The species involved in this case are $\{Fru6P, Fru16P2, Phi, ADPf, Pyrex\}$ and these therefore now appear in the edge labels (separators) of $\mathcal{T}_{\mathrm{MOD},2}$ rather than in the residuals. The proposition below establishes that the validity of the modularization remains unchanged by such an operation.

PROPOSITION 6.1. *Suppose that* $\{M_d | M_d \subseteq \mathcal{V}\}$, *is a modularization according to Definition 5.1 of a standard SKM* $[N, S, \mathsf{P}]$, *and that the modularization satisfies, for all* d, *the separation* $\{M_d\} \perp_{G^\sim} \{\bigcup_{e \neq d} M_e\} | S_d$ *in the undirected KIG* $G^\sim$. *Define a new collection of subsets* $\{\tilde{M}_d | \tilde{M}_d \subseteq \mathcal{V}\}$ *where* $\tilde{M}_d = M_d \cup \{\bigcup_{e \neq d} c_{ed}\}$ *and,* $\forall e \neq d$, $c_{ed} \subset M_e$ *and* $c_{ed} \cap M_d = \varnothing$ ($c_{ed} = \varnothing$ *being allowed*). *The species* c_{ed} *are called those "copied from e to d."*

Then $\{\tilde{M}_d\} \perp_{G^\sim} \{\bigcup_{e \neq d} \tilde{M}_e\} | \tilde{S}_d$ $\forall d$ *and, provided that Condition 4.3 continues to hold for* $\tilde{\Gamma}_d = \Delta(\tilde{M}_d \setminus \tilde{S}_d) \cap \Delta(\mathcal{V} \setminus \tilde{M}_d)$ $\forall d$, $\{\tilde{M}_d | \tilde{M}_d \subseteq \mathcal{V}\}$ *is also a modularization of the SKM* $[N, S, \mathsf{P}]$.

PROOF. Clearly, $\bigcup_d \tilde{M}_d = \mathcal{V}$. By Corollary 4.6, it suffices to show that $\{\tilde{M}_d\} \perp_{G^\sim} \{\bigcup_{e \neq d} \tilde{M}_e\} | \tilde{S}_d$ $\forall d$. Let $t_d := \bigcup_{e \neq d} c_{ed}$, the species copied to d, and $f_d := \bigcup_{e \neq d} c_{de}$, the species copied from d. The separation $\{M_d\} \perp_{G^\sim} \{\bigcup_{e \neq d} M_e\} | S_d$ implies that $\{M_d \cup t_d\} \perp_{G^\sim} \{\bigcup_{e \neq d} M_e\} \cup f_d | \{S_d \cup t_d \cup f_d\}$, which yields the required result since $\tilde{S}_d = \{M_d \cup t_d\} \cap [\{\bigcup_{e \neq d} M_e\} \cup f_d] = S_d \cup t_d \cup f_d \cup \varnothing$. □

There are 3 modules comprising $\mathcal{T}_{\mathrm{MOD},2}$ which together contain 45 different species, of which 32 distinct species are found only in module residuals (and hence are found in exactly 1 residual). The redox module $\{NADPHf, \ldots\}$ has already been discussed above. The module $\{Glcin, \ldots\}$ has the largest intersection with the rest of the network and acts as a linking module; it will be termed the *MPS* (*Monosaccharide-Phosphate Sugar* module). The two modules $\{NADPHf, \ldots\}$ and $\{Gri23P2f, \ldots\}$, by contrast, have only 2 species in common, namely $(NADH, Pyr)$—these are the only species common to all three modules. The *MPS* residual contains species that all belong to a single chemical class of molecule, namely monosaccharide sugar molecules (mostly with phosphate groups attached), with a further 6 different monosaccharide-phosphates (*MPs*) found in the rest of the module. Interestingly, the *MPs* of the module are those found in two "pathways" traditionally discussed separately—the pentose phosphate and

glycolytic pathways. Indeed, the *MPs* (*Glc6P*, *Fru6P*, *GraP*) all participate in reactions found in both "pathways."

The third and final module {*Gri23P2f*, ...} will be termed the *PGA* (*PhosphoGlycerate-Adenosine*) module, according to the chemical class of some of its constituents. It contains all of the phosphoglycerate molecules in the species set $\mathcal{V}$, namely (*Gri23P2f*, *MgGri23P2f*, *Gri13P2*, *Gri3P*, *Gri2P*), together also with all of the adenosine phosphate molecules (*ATPf*, *ADPf*, *AMPf*)—both free and complexed with magnesium (*Mg*). The module also contains all of the species involved in reactions of the so-called "pay-off phase" of glycolysis whose function is the production of the high-energy compounds *ATP* and *NADH*. That the dynamic evolution of, for example, all phosphoglycerates together with *PEP* is globally independent of all the other species in $\mathcal{V}$ conditional on the internal history of (*Fru16P2*, *MgATP*, *MgADP*, *GraP*, *NADH*, *Pyr*, *Rib5P*, *ADPf*, *Phi*) is again an insight provided by the modularization.

The modularizations $\mathcal{T}_{\mathrm{MOD},1}$ and $\mathcal{T}_{\mathrm{MOD},2}$ identified using the theory and methods developed in the paper constitute parsimonious, coarse-grained views of the metabolite network studied and provide important insight concerning the dynamics of the biological system as a whole.

7. Directions for future research. Application of the methods developed here to SKMs with large species sets and many component reactions is of considerable interest. In ongoing research that examines biochemical signalling networks with approximately 900 reactions and 750 species, the methods have been found to work effectively and to provide scientifically interesting modularizations.

It would be useful to consider methods for testing the adequacy of an SKM (perhaps augmented to allow for measurement error) as a statistical model of a given cellular system. Testing conditional independence relationships implied by a modularization of the SKM (such as the one in Figure 4 for the red blood cell) offers a promising means of assessing model adequacy. Clearly, it is not necessary to measure experimentally all species in the SKM, but all species in the relevant conditioning set (separator) must be measured. Intuitively, with $A \perp_{G^\sim} B|D$, changes in B—perhaps resulting from direct intervention on the levels of B—should be uninformative about changes in A over time intervals sufficiently short to ensure that levels of D usually remain constant (and vice versa).

SKMs subject to interventions are likely to become an area of active research, given their relevance both to medical and biotechnological applications. The predicted effect of interventions (e.g., gene knock-outs, RNA silencing, or receptor inhibition) could be derived by altering the specification of the SKM accordingly and comparing with the original SKM. There are also interesting connections with the causal inference literature more generally. Recently, Commenges and Gégout-Petit [2] introduced a "general dynamical model as a framework for causal interpretation," adopting an approach to causality based on "physical laws in sufficiently large systems." Local independence plays an important role in their analysis and

definition of influence. One might imagine that a sufficiently large SKM would be a candidate "perfect system" for a given smaller and observable cellular system. However, the jump processes followed by biochemical species and hence also SKMs do not belong to the class $(\mathcal{D})$ of special semimartingales to which [2] confines attention. Nevertheless, the approach seems relevant in broad terms. Finally, the experimental design of interventions to test causal claims derived from SKMs merits attention.

APPENDIX A: PROOFS FOR GLOBAL DYNAMIC INDEPENDENCE

A.1. Proof of Theorem 4.4. First we show that $\mathcal{F}_t^{\Delta(A)\cap\Delta(B)} = \mathcal{F}_{AB}^{D}(t) \subseteq \mathcal{F}_t^{D^*}$, in order to establish (A.1) below—that is, the internal history of all reactions that change A and B is contained in the internal history of $N^{D^*}(t)$.

The separation $A \perp_{G^\sim} B|D$ implies that for any $m \in \Delta(A) \cap \Delta(B)$, $R[m] \subseteq D$ [suppose not—then in the KIG G, either $\mathrm{pa}(B) \cap A \neq \varnothing$ or $\mathrm{pa}(A) \cap B \neq \varnothing$ which contradicts the separation]. For any $m \in \Delta(A) \cap \Delta(B)$, $R[m] \neq \varnothing$ and $R^*[m] \neq \varnothing$ by Definition 4.3(iii) and (iv); clearly $R^*[m] \subseteq D$. Hence $m \in \Delta(D)$ and $\Delta D_{AB} = \Delta(A) \cap \Delta(B)$ (with the possibility $\Delta D_{AB} = \varnothing$ not excluded). By Condition 4.3, any reaction in $\Gamma = \Delta(A) \cap \Delta(B)$ changes D differently—that is, the partition $\mathcal{M}(\Delta D_{AB})$ is either empty or consists of singletons—since $\forall m, \tilde{m} \in \Gamma\ (m \neq \tilde{m})$, $S_m^- \neq S_{\tilde{m}}^-$ and $(S_m^-)^A = (S_m^-)^B = \mathbf{0}$ because $R^*[m] \subseteq D$ (similarly for $\tilde{m}$), hence $(S_m^-)^D \neq (S_{\tilde{m}}^-)^D$ and $S_m^D \neq S_{\tilde{m}}^D$. Hence $N_t^{\Delta(A)\cap\Delta(B)} = N_{AB}^D(t)$ $\forall t$ and $\mathcal{F}_t^{\Delta(A)\cap\Delta(B)} = \mathcal{F}_{AB}^D(t) \subseteq \mathcal{F}_t^{D^*}$, which implies immediately that

$$\mathcal{F}_t^{\Delta(A)} \perp\!\!\!\perp \mathcal{F}_t^{\Delta(A)\cap\Delta(B)} | \mathcal{F}_t^{D^*}; \tilde{\mathsf{P}}_t. \tag{A.1}$$

Together with

$$\mathcal{F}_t^{\Delta(A)} \perp\!\!\!\perp \mathcal{F}_t^{\Delta(B)\setminus\Delta(A)} | \mathcal{F}_t^{D^*}; \tilde{\mathsf{P}}_t \tag{A.2}$$

(which is proved below) it follows that

$$\mathcal{F}_t^{\Delta(A)} \perp\!\!\!\perp \mathcal{F}_t^{\Delta(B)} | \mathcal{F}_t^{D^*}; \tilde{\mathsf{P}}_t,$$

since $\mathcal{F}_t^{\Delta(A)\cap\Delta(B)} \vee \mathcal{F}_t^{D^*} = \mathcal{F}_t^{D^*}$ and $\mathcal{F}_t^{\Delta(B)\setminus\Delta(A)} \vee \mathcal{F}_t^{\Delta(A)\cap\Delta(B)} = \mathcal{F}_t^{\Delta(B)}$. It then follows that $\mathcal{F}_t^A \perp\!\!\!\perp \mathcal{F}_t^B | \mathcal{F}_t^{D^*}; \tilde{\mathsf{P}}_t$ as required since it is clear from the definition of $N^A(t)$ and $N^B(t)$ that $\mathcal{F}_t^A \subseteq \mathcal{F}_t^{\Delta(A)}$ and $\mathcal{F}_t^B \subseteq \mathcal{F}_t^{\Delta(B)}$. (The reader unfamiliar with conditional independence of σ-fields and its properties is referred to [7]—see, in the context of this proof, Theorem 2.2.1, Corollary 2.2.4, Theorem 2.2.10 and Corollary 2.2.11 there.)

It remains to establish (A.2). Under $\tilde{\mathsf{P}}$, and hence also under $\tilde{\mathsf{P}}_t$, $\{\mathcal{F}_t^m | m = 1, \ldots, M\}$ are independent σ-fields (see Lemma 4.2). It follows that $\mathcal{F}_t^{\Delta(A)} \perp\!\!\!\perp \mathcal{F}_t^{\Delta(B)\setminus\Delta(A)} | \mathcal{F}_t^{\Delta D_D}; \tilde{\mathsf{P}}_t$ since $[\Delta(A), \Delta(B) \setminus \Delta(A), \Delta D_D]$ is a partition of $\{1, \ldots, M\}$. It now suffices for (A.2) to show the existence of $\mathcal{G}_t^{\Delta(A)} \subseteq \mathcal{F}_t^{\Delta(A)}$

and $\mathcal{G}_t^{\Delta(B)\setminus\Delta(A)} \subseteq \mathcal{F}_t^{\Delta(B)\setminus\Delta(A)}$, such that $\mathcal{G}_t^{\Delta(A)} \vee \mathcal{G}_t^{\Delta(B)\setminus\Delta(A)} \vee \mathcal{F}_t^{\Delta D_D} = \mathcal{F}_t^{D^*}$. Heuristically, we want to identify "information" contained only in $\mathcal{F}_t^{\Delta(A)}$ and $\mathcal{F}_t^{\Delta(B)\setminus\Delta(A)}$, respectively, which when jointly combined with $\mathcal{F}_t^{\Delta D_D}$ gives the internal history of $N_t^{D^*}$. But this corresponds exactly to the way $N_t^{D^*} = [\{N_A^D(t), N_{AB}^D(t)\}, N_B^D(t), N_D^D(t)]$ was constructed.

Recall Definition 4.2 for $N^{D^*}(t)$; its history $\mathcal{F}_t^{D^*}$ is given by $[\mathcal{F}_A^D(t) \vee \mathcal{F}_{AB}^D(t) \vee \mathcal{F}_B^D(t) \vee \mathcal{F}_D^D(t)]$. Since $\Delta D_A \subseteq \Delta(A)$, $\mathcal{F}_A^D(t) \subseteq \mathcal{F}_t^{\Delta D_A} \subseteq \mathcal{F}_t^{\Delta(A)}$; similarly $\Delta D_{AB} \subseteq \Delta(A)$ and hence $\mathcal{F}_{AB}^D(t) = \mathcal{F}_t^{\Delta D_{AB}} \subseteq \mathcal{F}_t^{\Delta(A)}$; and $\Delta D_B \subseteq \Delta(B) \setminus \Delta(A)$ hence $\mathcal{F}_B^D(t) \subseteq \mathcal{F}_t^{\Delta D_B} \subseteq \mathcal{F}_t^{\Delta(B)\setminus\Delta(A)}$. Note that $\mathcal{F}_D^D(t) = \mathcal{F}_t^{\Delta D_D}$ since $\mathcal{M}(\Delta D_D)$ is either empty or consists of singletons—any 2 reactions that change D alone must do so differently since no 2 columns of S are equal (by Definition 2.1). Finally, taking $\mathcal{G}_t^{\Delta(A)} = \mathcal{F}_A^D(t) \vee \mathcal{F}_{AB}^D(t) \subseteq \mathcal{F}_t^{\Delta(A)}$ and $\mathcal{G}_t^{\Delta(B)\setminus\Delta(A)} = \mathcal{F}_B^D(t) \subseteq \mathcal{F}_t^{\Delta(B)\setminus\Delta(A)}$ completes the proof since then $\mathcal{G}_t^{\Delta(A)} \vee \mathcal{G}_t^{\Delta(B)\setminus\Delta(A)} \vee \mathcal{F}_t^{\Delta D_D} = \mathcal{F}_t^{D^*}$ as required.

A.2. Proof of Theorem 4.5. The proof is in 3 parts. (I) First show that $\forall m \in \mathcal{M}$, either $R[m] \subseteq A \cup D$ in which case $\int_0^t \lambda_m(u)\,du$ is adapted to $\mathcal{F}_t^{AD^*}$, or $R[m] \subseteq B \cup D$ in which case $\int_0^t \lambda_m(u)\,du$ is adapted to $\mathcal{F}_t^{BD^*}$.

The separation $A \perp_{G^\sim} B|D$ implies that either $R[m] \subseteq A \cup D$ or $R[m] \subseteq B \cup D$. Suppose not, then $B \cap R[m] \neq \varnothing$ and $A \cap R[m] \neq \varnothing$—arguing using (i) of Definition 4.3, either $m \in \Delta(A)$ in which case $B \cap \mathrm{pa}(A) \neq \varnothing$, which contradicts the separation; or $m \in \Delta(B)$ in which case $A \cap \mathrm{pa}(B) \neq \varnothing$, which also contradicts the separation. If $B \cap R[m] \neq \varnothing$ and $A \cap R[m] \neq \varnothing$, then $m \in \Delta D_D$ is not possible—if $m \in \Delta D_D$ then the reactants that are changed $R^*[m] \subseteq D$ and hence, by (iv) of Definition 4.3, either $B \cap R[m] \neq \varnothing$ or$A \cap R[m] \neq \varnothing$ but not both.

Therefore, if $R[m] \subseteq A \cup D$ (resp., $R[m] \subseteq B \cup D$) then both $\lambda_m(t)$ and $\log(\lambda_m(t))$ are measurable with respect to $\mathcal{F}_t^{R[m]} \subseteq \mathcal{F}_t^{A\cup D} \subseteq \mathcal{F}_t^{AD^*}$ (resp., $\mathcal{F}_t^{B\cup D} \subseteq \mathcal{F}_t^{BD^*}$) by (2.6), since $X^{R[m]}(t-)$ is measurable $\mathcal{F}_t^{R[m]}$ and $\mathcal{F}_t^D \subseteq \mathcal{F}_t^{D^*}$. Since $\lambda_m(t)$ is also càglàd, if $R[m] \subseteq A \cup D$ (resp., $R[m] \subseteq B \cup D$) then $\lambda_m(t)$ is $\mathcal{F}_t^{AD^*}$-predictable and hence $\int_0^t \lambda_m(u)\,du$ is $\mathcal{F}_t^{AD^*}$-adapted (resp., $\mathcal{F}_t^{BD^*}$-adapted).

(II) Second show that if $R[m] \subseteq A \cup D$ (resp., $R[m] \subseteq B \cup D$), then $\{T_s^m\}_{s\geq 1}$ are $\mathcal{F}_t^{AD^*}$-stopping times (resp., $\mathcal{F}_t^{BD^*}$-stopping times). It will then follow that $1(T_s^m \leq t)\log(\lambda_m(T_s^m))$ is $\mathcal{F}_t^{AD^*}$-measurable $\forall s \geq 1$ (resp., $\mathcal{F}_t^{BD^*}$-measurable) by the definition of $\mathcal{F}_{T_s^m}^{AD^*}$, because $\log(\lambda_m(t))$ is left continuous and hence $\log(\lambda_m(T_s^m))$ is $\mathcal{F}_{T_s^m}^{AD^*}$-measurable—see, for example, Theorem 2.1.10 of [20]. This in turn yields that $\sum_{s\geq 1} 1(T_s^m \leq t)\log(\lambda_m(T_s^m))$ is $\mathcal{F}_t^{AD^*}$-measurable (resp., $\mathcal{F}_t^{BD^*}$-measurable). To establish the required stopping time property for $\{T_s^m\}_{s\geq 1}$, distinguish the following cases, exactly one of which must hold $\forall m \in \mathcal{M}$:

(i) $m \in \Delta D_D$: recall that $\mathcal{M}(\Delta D_D)$ consists of singletons since any 2 reactions that change D alone must do so differently (by Definition 2.1). Therefore $N_m(t)$

is adapted to $\mathcal{F}_t^{\Delta D_D} = \mathcal{F}_D^D(t) \subseteq \mathcal{F}_t^{D^*}$, hence $\{T_s^m\}_{s\geq 1}$ are $\mathcal{F}_t^{D^*}$-stopping times. Either $R[m] \subseteq A \cup D$ or $R[m] \subseteq B \cup D$ [by part (I) above]. If $R[m] \subseteq A \cup D$ (resp., $R[m] \subseteq B \cup D$), then $\{T_s^m\}_{s\geq 1}$ are necessarily $\mathcal{F}_t^{AD^*}$-stopping times (resp., $\mathcal{F}_t^{BD^*}$-stopping times), as required.

(ii) $m \in \Delta A \cap \Delta(B)$: recall from the proof of Theorem 4.4 that $R[m] \subseteq D$ and the partition $\mathcal{M}(\Delta D_{AB})$ consists of singletons. Hence $\mathcal{F}_t^m \subseteq \mathcal{F}_t^{\Delta A \cap \Delta(B)} = \mathcal{F}_{AB}^D(t) \subseteq \mathcal{F}_t^{D^*}$, $N_m(t)$ is adapted to $\mathcal{F}_t^{D^*}$ and $\{T_s^m\}_{s\geq 1}$ are $\mathcal{F}_t^{D^*}$-stopping times.

(iii) $m \in \Delta(A) \setminus \Delta(B)$: we have that $R[m] \subseteq A \cup D$ by part (I) above; consider the cases (iiia) $m \in \Delta D_A$, and (iiib) $m \notin \Delta D_A$ in turn below to conclude that in each case $\{T_s^m\}_{s\geq 1}$ are $\mathcal{F}_t^{AD^*}$-stopping times.

(iiia) Identify the element of $\mathcal{M}(\Delta D_A)$ corresponding to changes in D equal to S_m^D, $\mathcal{M}_e(\Delta D_A)$ say. Denote the corresponding element of $N_A^D(t)$ by $N_{A,e}^D(t)$, which is clearly measurable $\mathcal{F}_t^{D^*}$, and the jump times of this univariate counting process by $\{T_{A,e}^D(s)\}_{s\geq 1}$. Thus $\{T_{A,e}^D(s)\}_{s\geq 1}$ are $\mathcal{F}_t^{AD^*}$-stopping times. Now $\mathcal{M}_e(\Delta D_A)$ may not be a singleton, but we can write

$$N_m(t) = \sum_{s\geq 1} 1\{T_{A,e}^D(s) \leq t\} 1\{X^{A\cup D}(T_{A,e}^D(s)) - X^{A\cup D}(T_{A,e}^D(s)-) = S_m^{A\cup D}\},$$

since $\nexists \tilde{m} \in \mathcal{M}_e(\Delta D_A)$ $(m \neq \tilde{m})$ s.t. $S_m^{A\cup D} = S_{\tilde{m}}^{A\cup D}$ (by Definition 2.1 and $S_m^B = S_{\tilde{m}}^B = \mathbf{0}$). Since $X^{A\cup D}(t)$ is right continuous and $X^{A\cup D}(t-)$ left continuous, and both are $\mathcal{F}_t^{AD^*}$-adapted since $\mathcal{F}_t^{A\cup D}$-adapted, $[X^{A\cup D}(T_{A,e}^D(s)) - X^{A\cup D}(T_{A,e}^D(s)-)]$ is $\mathcal{F}^{AD^*}(T_{A,e}^D(s))$-measurable (by, e.g., Theorem 2.1.10 of [20]). Hence the summand is $\mathcal{F}^{AD^*}(t)$-measurable $\forall s \geq 1$, $N_m(t)$ is adapted to $\mathcal{F}_t^{AD^*}$ and $\{T_s^m\}_{s\geq 1}$ are $\mathcal{F}_t^{AD^*}$-stopping times.

(iiib) Then $m \in \Delta^*(A) := \Delta(A) \setminus (\Delta(B) \cup \Delta(D))$. Define for any p-variate counting process $N(t)$ $(p \geq 1)$, the "ground process" $\bar{N}(t) := \mathbf{1}'_{p\times 1} N(t)$. We may then write

$$\bar{N}^{\Delta^*(A)}(t) = \bar{N}^A(t) - \bar{N}_A^D(t) - \bar{N}_{AB}^D(t),$$

where $\bar{N}^{\Delta^*(A)}(t)$ is the number of reactions on $[0,t]$ that change A alone [noting that $\Delta(A) \cap \Delta(B) = \Delta D_{AB}$]. Hence $\bar{N}^{\Delta^*(A)}(t)$ is measurable $\mathcal{F}_t^A \vee \mathcal{F}_A^D(t) \vee \mathcal{F}_{AB}^D(t) \subseteq \mathcal{F}_t^{AD^*}$, and its jump times $\{T_s^{\Delta^*(A)})\}_{s\geq 1}$ are $\mathcal{F}_t^{AD^*}$-stopping times. Also,

$$N_m(t) = \sum_{s\geq 1} 1\{T_s^{\Delta^*(A)} \leq t\} 1\{X^A(T_s^{\Delta^*(A)}) - X^A(T_s^{\Delta^*(A)}-) = S_m^A\},$$

since $\nexists \tilde{m} \in \Delta^*(A)$ $(m \neq \tilde{m})$ s.t. $S_m^A = S_{\tilde{m}}^A$ (by Definition 2.1 and $S_m^{B\cup D} = S_{\tilde{m}}^{B\cup D} = \mathbf{0}$). Since $X^A(t)$ is right continuous and $X^A(t-)$ left continuous, and both are $\mathcal{F}_t^{AD^*}$-adapted since $\mathcal{F}_t^A$-adapted, $[X^A(T_s^{\Delta^*(A)}) - X^A(T_s^{\Delta^*(A)}-)]$ is $\mathcal{F}^{AD^*}(T_s^{\Delta^*(A)})$-measurable. Hence the summand is $\mathcal{F}^{AD^*}(t)$-measurable $\forall s \geq 1$, and $\{T_s^m\}_{s\geq 1}$ are $\mathcal{F}_t^{AD^*}$-stopping times.

(iv) $m \in \Delta(B) \setminus \Delta(A)$: we have that $R[m] \subseteq B \cup D$ by part (I) above; argue as in (iii) with A in place of B and vice versa to conclude that $\{T_s^m\}_{s\geq 1}$ are $\mathcal{F}_t^{BD^*}$-stopping times.

(III) Combining parts (I) and (II) above establishes that if $R[m] \subseteq A \cup D$ (resp., $R[m] \subseteq B \cup D$) then $\mathcal{L}_{m,t} := \exp(l_m(t))$ is measurable $\mathcal{F}_t^{AD^*}$ (resp., $\mathcal{F}_t^{BD^*}$). Then, in an obvious manner, grouping the $\mathcal{L}_{m,t}$ into 2 groups according to the forementioned measurability property and defining the $\psi_{iD^*,t}$ as the product within each group yields $\mathcal{L}_t = \psi_{AD^*,t} \cdot \psi_{BD^*,t}$, where $\psi_{iD^*,t}$ is nonnegative and $\mathcal{F}_t^i \vee \mathcal{F}_t^{D^*}$-measurable for $i \in \{A, B\}$.

APPENDIX B: ADDITIONAL PROOFS

PROOF OF LEMMA 2.1. It remains to establish that $\mathcal{F}_t^A = \mathcal{F}_t^{X^A}$. First show that $\mathcal{F}_t^A \supseteq \mathcal{F}_t^{X^A}$. We have that $\mathcal{F}_t^A = \sigma(Z_s^A 1(T_s^A \leq u); 0 \leq u \leq t,\ s \geq 1)$ and $X^A(u) = X^A(0) + \sum_{s\geq 1} Z_s^A 1(T_s^A \leq u)$, which is therefore measurable $\mathcal{F}_t^A$. Second show that $\mathcal{F}_t^A \subseteq \mathcal{F}_t^{X^A}$. We have also that $\mathcal{F}_t^A = \sigma(1(Z_s^A = S_m^A)1(T_s^A \leq u); 0 \leq u \leq t,\ s \geq 1, m \in \Delta A)$, hence it suffices to show that $1(Z_s^A = S_m^A)1(T_s^A \leq u)$ is measurable $\mathcal{F}_t^{X_A}$. By its construction, $\{T_s^A\}$ are the jump times of the right-continuous jump process X^A. The filtration $\{\mathcal{F}_t^{X_A}\}$ is right continuous. Hence, for $s \geq 1$, T_s^A is an $\mathcal{F}_t^{X_A}$-stopping time and $X^A(T_s^A)$ is $\mathcal{F}^{X_A}(T_s^A)$-measurable. Since $Z_s^A = X^A(T_s^A) - X^A(T_{s-1}^A)$ and $\mathcal{F}^{X_A}(T_{s-1}^A) \subseteq \mathcal{F}^{X_A}(T_s^A)$, Z_s^A is also $\mathcal{F}^{X_A}(T_s^A)$-measurable. Hence $\{1(Z_s^A = S_m^A) = 1\} \cap \{1(T_s^A \leq u) = 1\} \in \mathcal{F}^{X_A}(u) \subseteq \mathcal{F}^{X_A}(t)$ by the definition of $\mathcal{F}^{X_A}(T_s^A)$, and therefore $1(Z_s^A = S_m^A)1(T_s^A \leq u)$ is measurable $\mathcal{F}_t^{X_A}$. □

PROOF OF LEMMA 4.1. Let $\mathcal{L}_{i3} := (\mathsf{dP}/\mathsf{d}\tilde{\mathsf{P}})|_{\mathcal{F}^i \vee \mathcal{F}^3}$, and $\mathcal{L}_3 := (\mathsf{dP}/\mathsf{d}\tilde{\mathsf{P}})|_{\mathcal{F}^3}$. Then it is straightforward to show that $\mathcal{L}_{i3} = \tilde{\mathsf{E}}[\mathcal{L}_{123}|\mathcal{F}^i \vee \mathcal{F}^3]$ and $\mathcal{L}_3 = \tilde{\mathsf{E}}[\mathcal{L}_{123}|\mathcal{F}^3]$, where $\tilde{\mathsf{E}}$ denotes expectation under $\tilde{\mathsf{P}}$. Hence, $\mathcal{L}_{13} = \psi_{13}\tilde{\mathsf{E}}[\psi_{23}|\mathcal{F}^1 \vee \mathcal{F}^3]$ and $\mathcal{L}_{23} = \psi_{23}\tilde{\mathsf{E}}[\psi_{13}|\mathcal{F}^1 \vee \mathcal{F}^3]$ by the nonnegativity and measurability of the ψ_{i3}. Since $\mathcal{F}^2 \vee \mathcal{F}^3 \perp\!\!\!\perp \mathcal{F}^1|\mathcal{F}^3; \tilde{\mathsf{P}}$, $\tilde{\mathsf{E}}[\psi_{23}|\mathcal{F}^1 \vee \mathcal{F}^3] = \tilde{\mathsf{E}}[\psi_{23}|\mathcal{F}^3]$ by Definition 4.1 and hence $\mathcal{L}_{13} = \psi_{13}\tilde{\mathsf{E}}[\psi_{23}|\mathcal{F}^3]$. Similarly, $\mathcal{L}_{23} = \psi_{23}\tilde{\mathsf{E}}[\psi_{13}|\mathcal{F}^3]$. Furthermore, $\mathcal{L}_3 = \tilde{\mathsf{E}}[\psi_{13}|\mathcal{F}^3]\tilde{\mathsf{E}}[\psi_{23}|\mathcal{F}^3]$ by the nonnegativity and measurability of the ψ_{i3} and since $\mathcal{F}^1 \vee \mathcal{F}^3 \perp\!\!\!\perp \mathcal{F}^2 \vee \mathcal{F}^3|\mathcal{F}^3; \tilde{\mathsf{P}}$. Therefore,

$$\mathcal{L}_{123}\mathcal{L}_3 = \mathcal{L}_{13}\mathcal{L}_{23} \tag{B.1}$$

and, in particular, $\mathcal{L}_{123}\mathcal{L}_3 = \mathcal{L}_{13}\mathcal{L}_{23}$ on the event $\{\mathcal{L}_3 = \tilde{\mathsf{E}}[\psi_{13}|\mathcal{F}^3]\tilde{\mathsf{E}}[\psi_{23}|\mathcal{F}^3] > 0\}$, whence $\mathcal{F}^1 \perp\!\!\!\perp \mathcal{F}^2|\mathcal{F}^3; \mathsf{P}$ by Theorem 2.2.14 of [7]. □

PROOF OF PROPOSITION 5.2. The proof is in 3 steps, according to the number of pairs of clusters aggregated under step 5 of Algorithm 5.1: (i) for the case where no pair of clusters is aggregated, and hence $\mathcal{T}_{\mathrm{MOD}} = \mathcal{T}_C$; (ii) for the case where exactly 1 pair of clusters is aggregated; (iii) for the case where more than 1 pair of clusters is aggregated.

(i) $\mathcal{T}_C$ is a junction tree representation of the clique decomposition of $G_T^{\sim}$. For the proof of this case see the proof of Theorem 4.6 of [3].

(ii) $\mathcal{T}_{\text{MOD}}$ is connected (as a consequence of $\mathcal{T}_C$ being connected), and has $(\delta - 1)$ nodes and $(\delta - 2)$ edges (one less edge than $\mathcal{T}_C$); $\mathcal{T}_{\text{MOD}}$ is therefore a tree, whence there is a unique path in $\mathcal{T}_{\text{MOD}}$ between any pair (M_d, M_e) of its clusters. It is straightforward (but somewhat tedious) to show that every cluster on this path must contain $M_d \cap M_e$ since the corresponding path in $\mathcal{T}_C$ possesses this junction property [by (i) above]. Hence $\mathcal{T}_{\text{MOD}}$ is a junction tree. It remains to prove that for any 2 adjacent clusters (M_d, M_e) in $\mathcal{T}_{\text{MOD}}$, we have $M_d \cap M_e = V_{de} \cap V_{ed}$ and $V_{de} \perp_{G_T^{\sim}} V_{ed} | S_{de}$.

We will show (iia) that edges "in common" between $\mathcal{T}_C$ and $\mathcal{T}_{\text{MOD}}$—the $(\delta - 2)$ edges not removed by the cluster aggregation—carry the same label, that is, the intersection of the clusters joined by each such edge is unchanged; and (iib) that cutting any such edge in both $\mathcal{T}_C$ and $\mathcal{T}_{\text{MOD}}$ results in pairs of subtrees whose clusters have identical unions in the two cases. The result then follows from (i) above.

(iia) If both clusters, (M_d, M_e), joined by such an edge, are in $\mathcal{T}_C$ and $\mathcal{T}_{\text{MOD}}$ the claim is obviously true. Consider then the case where M_d, say, is the result of the aggregation of the cluster pair (M_α, M_β). Suppose, without loss of generality, that $M_\alpha \sim M_e$ in $\mathcal{T}_C$. Now $S_{de} = (M_e \cap M_\alpha) \cup (M_e \cap M_\beta)$. The edge joining M_d to M_e in $\mathcal{T}_{\text{MOD}}$ was formerly, in $\mathcal{T}_C$, the edge $M_\alpha \sim M_e$, whence $S_{de} = (M_e \cap M_\alpha)$ since M_α is on the path between M_β and M_e in $\mathcal{T}_C$ and $(M_e \cap M_\beta) \subseteq M_\alpha$. Thus, the intersection of the clusters joined by the edge is always the same in $\mathcal{T}_{\text{MOD}}$ and $\mathcal{T}_C$, as claimed.

(iib) Let the edge that is cut in both cases be $M_d \sim M_e$ [where it is understood that M_d, say, may be equal to M_α in $\mathcal{T}_C$ and hence equal to $(M_\alpha \cup M_\beta)$ in $\mathcal{T}_{\text{MOD}}$]. It is required to show, using an obvious notation, that $V_{de}^{\text{MOD}} = V_{de}^{C}$ and $V_{ed}^{\text{MOD}} = V_{ed}^{C}$. It is well known that cutting an edge in any tree results in 2 disconnected subtrees. One of the 2 pairs of subtrees generated here must contain 2 identical subtrees. Suppose then, without loss of generality, that $\mathcal{T}_{\text{MOD}}^{ed} = \mathcal{T}_C^{ed}$, whence $V_{ed}^{\text{MOD}} = V_{ed}^{C}$. The subtrees $\mathcal{T}_{\text{MOD}}^{de}$ and $\mathcal{T}_C^{de}$ have the same clusters, except for the aggregation of the cluster pair (M_α, M_β) to form $M_{\alpha\beta}$ in $\mathcal{T}_{\text{MOD}}^{de}$. It is straightforward (but tedious) to show that, for $\gamma \notin \{\alpha, \beta\}$ and M_γ a cluster in $\mathcal{T}_C^{de}$,

$$[M_\gamma]_{\mathcal{T}_C^{de}} \setminus \{M_\alpha, M_\beta\} = [M_\gamma]_{\mathcal{T}_{\text{MOD}}^{de}} \setminus \{M_{\alpha\beta}\},$$

where $[M]_{\mathcal{T}}$ are the clusters that can be reached from cluster M by paths in a tree $\mathcal{T}$. The subtrees $\mathcal{T}_\bullet^{de}$ are themselves connected graphs, but disconnected from the corresponding $\mathcal{T}_\bullet^{ed}$. Therefore the clusters of $\mathcal{T}_\bullet^{de}$ are given exactly by $[M_\gamma]_{\mathcal{T}_\bullet^{de}}$, where M_γ is any one of its clusters. It follows that

$$\begin{aligned} V_{de}^{C} &= \left\{ \bigcup [M_\gamma]_{\mathcal{T}_C^{de}} \setminus \{M_\alpha, M_\beta\} \right\} \cup M_\alpha \cup M_\beta \\ &= \left\{ \bigcup [M_\gamma]_{\mathcal{T}_{\text{MOD}}^{de}} \setminus \{M_{\alpha\beta}\} \right\} \cup M_{\alpha\beta} = V_{de}^{\text{MOD}} \end{aligned}$$

as required.

(iii) The proof is by induction on the number of cluster pairs, n say, that are aggregated. Parts (i) and (ii) above establish the proposition for $n = 0$ and $n = 1$. Exactly the same mode of argument as the one used in (ii) above also establishes that if the proposition holds for $n \geq 0$, it must hold for $(n + 1)$. This completes the proof. □

APPENDIX C: SPECIES NAMES FOR RED BLOOD CELL SKM

AMPf = AMP (unbound); *ADPf* = ADP; *ATPf* = ATP; *DHAP* = Dihydroxyacetone phosphate; *E4P* = Erythrose 4-phosphate; *Fru6P* = Fructose 6-phosphate; *Fru16P2* = Fructose 1,6-phosphate; *GlcA6P* = Phospho-D-glucono-1,5-lactone; *Glcin* = Glucose (cytoplasmic); *Glcout* = External Glucose; *Glc6P* = Glucose 6-phosphate; *GraP* = Glyceraldehyde 3-phosphate; *Gri13P2* = 1,3-Bisphospho-D-glycerate; *Gri3P* = 3-Phospho-D-glycerate; *Gri23P2* = 2,3-Bisphospho-D-glycerate; *Gri2P* = 2-Phospho-D-glycerate; *GSH* = Reduced Glutathione; *GSSG* = Oxidized Glutathione; *Lac* = Lactate; *Lacex* = External Lactate; *MgATP*; *MgADP*; *MgAMP*; *Mg*; *MgGri23P2*; *NADH*; *NADPf* = NADP (unbound); *NADPHf* = NADPH; *P1f* = Protein1; *P2* = Protein2; *P1NADP* = Protein1 bound NADP; *P1NADPH* = Protein1 bound NADPH; *P2NADP* = Protein2 bound NADP; *P2NADPH* = Protein2 bound NADPH; *PEP* = Phosphoenolpyruvate; *Phi* = Phosphate; *NAD*; *PRPP* = Phosphoribosylpyrophosphate; *Pyr* = Pyruvate; *Pyrex* = External Pyruvate; *Rib5P* = Ribose 5-phosphate; *Rul5P* = Ribulose 5-phosphate; *Sed7P* = Sedoheptulose 7-phosphate; *Xul5P* = Xylulose 5-phosphate.

Acknowledgments. The author is grateful for the research environment provided by the Statistical Laboratory and the Cambridge Statistics Initiative, and to A. P. Dawid, V. Didelez, C. Holmes, M. Jacobsen and D. J. Wilkinson for helpful discussions. The comments of the Associate Editor and referees were valuable in improving the paper. Computations were performed using R version 2.8.1 and R packages gRbase and Rgraphviz.

REFERENCES

[1] BALL, K., KURTZ, T. G., POPOVIC, L. and REMPALA, G. (2006). Asymptotic analysis of multiscale approximations to reaction networks. *Ann. Appl. Probab.* **16** 1925–1961. MR2288709

[2] COMMENGES, D. and GÉGOUT-PETIT, A. (2009). A general dynamical statistical model with causal interpretation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 719–736.

[3] COWELL, R. G., DAWID, A. P., LAURITZEN, S. L. and SPIEGELHALTER, D. J. (2007). *Probabilistic Networks and Expert Systems*. Springer, New York. MR1697175

[4] DELLACHERIE, C. and MEYER, P.-A. (2006). *Probabilities and Potential*. North-Holland, Amsterdam. MR0521810

[5] DIDELEZ, V. (2007). Graphical models for composable finite Markov processes. *Scand. J. Statist.* **34** 169–185. MR2325249

[6] DIDELEZ, V. (2008). Graphical models for marked point processes based on local independence. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 245–264. MR2412641
[7] FLORENS, J., MOUCHART, M. and ROLIN, J. (1990). *Elements of Bayesian Statistics*. CRC Press, New York.
[8] GILLESPIE, C. (2009). Moment-closure approximations for mass-action models. *IET Systems Biology* **3** 52–58.
[9] GILLESPIE, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22** 403–434. MR0503370
[10] GILLESPIE, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* **81** 2340–2361.
[11] GILLESPIE, D. T. (1992). A rigorous derivation of the chemical master equation. *Phys. A* **188** 404–425.
[12] GOUTSIAS, J. (2007). Classical versus stochastic kinetics modeling of biochemical reaction systems. *Biophysical Journal* **92** 2350–2365.
[13] GUIMERÀ, R. and AMARAL, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature* **433** 895–900.
[14] HOLZHUETTER, H. (2004). The principle of flux minimization and its application to estimate stationary fluxes in biochemical networks. *European Journal of Biochemistry* **271** 2905–2922.
[15] HUBER, W., CAREY, V. J., LONG, L., FALCON, S. and GENTLEMAN, R. (2007). Graphs in molecular biology. *BMC Bioinformatics* **8** S8.
[16] JACOBSEN, M. (1982). *Statistical Analysis of Counting Processes*. *Lecture Notes in Statistics* **12**. Springer, New York. MR0676128
[17] JACOBSEN, M. (2006). *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes*. Birkhäuser, Boston. MR2189574
[18] KASHTAN, N. and ALON, U. (2005). Spontaneous evolution of modularity and network motifs. *Proc. Natl. Acad. Sci.* **102** 13773–13778.
[19] KAUFFMAN, K., PAJEROWSKI, J., JAMSHIDI, N., PALSSON, B. Ø. and EDWARDS, J. (2002). Description and analysis of metabolic connectivity and dynamics in the human red blood cell. *Biophysical Journal* **83** 646–662.
[20] LAST, G. and BRANDT, A. (1995). *Marked Point Processes on the Real Line: The Dynamic Approach*. Springer, New York. MR1353912
[21] LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. Clarendon Press, Oxford. MR1419991
[22] NODELMAN, U., SHELTON, C. and KOLLER, D. (2002). Continuous time Bayesian networks. In *Proc. 18th Conference on Uncertainty in Artificial Intelligence* 378–387. Morgan Kaufmann, San Francisco.
[23] NOVÈRE, N. L. (2006). Biomodels database: Model number BIOMD0000000070. Available at http://www.ebi.ac.uk/biomodels-main/BIOMD0000000070.
[24] OLESEN, K. G. and MADSEN, A. L. (2002). Maximal prime subgraph decomposition of Bayesian networks. *IEEE Trans. Syst. Man Cyb. Part B* **32** 21–31.
[25] PRICE, N., REED, J., PAPIN, J., WIBACK, S. and PALSSON, B. Ø. (2003). Network-based analysis of metabolic regulation in the human red blood cell. *J. Theoret. Biol.* **225** 185–194.
[26] REED, J. L., FAMILI, I., THIELE, I. and PALSSON, B. O. (2006). Towards multidimensional genome annotation. *Nature Reviews Genetics* **7** 130–141.
[27] SCHUSTER, R. and HOLZHUETTER, H. (1995). Use of mathematical models for predicting metabolic effect of large-scale enzyme activity alterations. *European Journal of Biochemistry* **229** 403–418.
[28] SHAHREZAEI, V. and SWAIN, P. S. (2008). The stochastic nature of biochemical networks. *Current Opinion in Biotechnology* **19** 369–374.

[29] SZALLASI, Z., PERIWAL, V. and STELLING, J. (2006). *System Modeling in Cellular Biology*. MIT Press, Cambridge.
[30] WILKINSON, D. J. (2009). Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics* **10** 122–133.

CENTRE FOR MATHEMATICAL SCIENCES
UNIVERSITY OF CAMBRIDGE
WILBERFORCE ROAD, CAMBRIDGE
UNITED KINGDOM
E-MAIL: C.Bowsher@statslab.cam.ac.uk