



HAL
open science

Automatic network reconstruction from experimental time-series data: A survey

Annegret K Wagler

► **To cite this version:**

Annegret K Wagler. Automatic network reconstruction from experimental time-series data: A survey. Information Technology, 2014, 56 (2), 10.1515/itit-2013-1020 . hal-02047311

HAL Id: hal-02047311

<https://hal.science/hal-02047311>

Submitted on 24 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Automatische Netzwerk-Rekonstruktion experimenteller Zeitserien-Daten:
Ein Überblick**

**Automatic network reconstruction from experimental time-series data:
A survey**

Prof. Dr. Annegret Wagler: CNRS and Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes (LIMOS), Université Blaise Pascal (Clermont-Ferrand II), Campus scientifique des Cézeaux, Bat. ISIMA, 63173 Aubière, France

Tel: +33 (0) 473 40 5367, Fax: +33 (0) 473 40 5001, E-Mail: wagler@isima.fr

Annegret Wagler received the PhD Degree from the University of Technology Berlin in 2000. From 2000 to 2004, she was post-doc at the Zuse Institute Berlin. From 2004 to 2010, she was the head of a junior research group at the Otto-von-Guericke-University Magdeburg and habilitated in 2007. In September 2010, she got a professor position at the University Blaise Pascal (Clermont-Ferrand II) and a chair of excellence of CNRS.

Keywords: G.2.1 [Mathematics of Computing: Discrete Mathematics: Combinatorics]; G.2.3 [Mathematics of Computing: Discrete Mathematics: Applications]; G.1.6 [Mathematics of Computing: Numerical Analysis: Optimization]

Schlagworte: Netzwerk-Rekonstruktion; kombinatorische Algorithmen; diskrete Strukturen

MS-ID:

wagler@isima.fr

November 5, 2013

Heft: 53/* (2011)

Abstract

The reconstruction of models from experimental data is a challenging problem due to the inherited complexity of the studied biological systems. We discuss an exact, exclusively data-driven approach to reconstruct Petri nets, a framework which turned out to coherently model static interactions and dynamic processes. The reconstructed models shall reproduce the experimentally observed dynamic behavior in a simulation. For that, we consider Petri nets with two types of extensions, priority relations among the transitions and control-arcs, to obtain additional activation rules that control the dynamic behavior. Here, we give an overview on results concerning the reconstruction of standard networks with and without priorities, extended Petri nets, and extended Petri nets with priorities. All these approaches aim at reconstructing all networks of the studied type that fit the given experimental data, to provide all possible alternatives of mechanisms behind the experimentally observed phenomena.

Zusammenfassung

Die Rekonstruktion von Modellen ausgehend von experimentellen Daten ist aufgrund der Komplexität biologischer Systeme ein schwieriges Problem. Wir stellen ein exaktes, ausschließlich auf der Analyse der gegebenen Daten beruhendes Verfahren vor, um Petri-Netze zu rekonstruieren: Modelle, die sowohl statische Interaktionen als auch dynamische Prozesse abbilden können. Die rekonstruierten Modelle sollen dabei in einer Simulation das experimentell beobachtete dynamische Verhalten aufweisen. Dafür betrachten wir Petri-Netze mit zweierlei Erweiterungen, Prioritäten auf Transitionen und Testkanten, um zusätzliche Aktivierungsregeln zu erhalten, die das dynamische Verhalten kontrollieren. Wir geben einen Überblick über Resultate zur Rekonstruktion von Standard-Netzwerken mit und ohne Prioritäten sowie erweiterten Petri-Netzen mit und ohne Prioritäten. Alle diese Verfahren haben zum Ziel, alle Netzwerke des betrachteten Typs zu rekonstruieren, die konform mit den gegebenen experimentellen Daten sind, um alle möglichen Alternativen von Mechanismen aufzuzeigen, die den experimentell beobachteten Phänomenen zugrunde liegen können.

1 Introduction

The overall aim of systems biology is to analyze biological systems and to understand different phenomena therein as, e.g., responses of cells to environmental changes, host-pathogen interactions, or effects of gene defects. To gain the required insight into the underlying biological processes, experiments are performed and the resulting experimental data are interpreted in terms of models that reflect the observed phenomena.

Depending on the biological aim and the type and quality of the available data, different types of mathematical models are used and corresponding methods for their reconstruction have been developed, see for instance [1, 15, 16]. Our work is dedicated to Petri nets, a framework which turned out to coherently model both static interactions in terms of networks and dynamic processes in terms of state changes, see e.g. [2, 13, 14, 18, 19].

In fact, a network $\mathcal{P} = (P, T, A, w)$ reflects the involved system components by places $p \in P$ and their interactions by transitions $t \in T$, the arcs in $A \subset (P \times T) \cup (T \times P)$ link places and transitions, and the arc weights $w : A \rightarrow \mathbb{N}$ reflect stoichiometric coefficients of the corresponding reactions. Moreover, each place $p \in P$ can be marked with an integral number x_p of tokens defining a system state $\mathbf{x} \in \mathbb{Z}_+^{|P|}$; we denote by \mathcal{X} the set of all potential system states. A transition $t \in T$ is *enabled* in a state \mathbf{x} if $x_p \geq w(p, t)$ for all p with $(p, t) \in A$, switching or firing t leads to a successor state \mathbf{x}' with $x'_p = x_p - w(p, t)$ for all $(p, t) \in A$ and $x'_p = x_p + w(t, p)$ for all $(t, p) \in A$. Dynamic processes are represented by sequences of such state changes.

Our central question is to reconstruct models of this type from experimental time-series data by means of an exact, exclusively data-driven approach. This approach takes as input a set P of places and discrete time-series data \mathcal{X}' given by sequences $(\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^k)$ of experimentally observed system states. The goal is to determine all Petri nets (P, T, A, w) that are able to reproduce the data, i.e., where T contains enough transitions to perform for each $\mathbf{x}^j \in \mathcal{X}'$ the experimentally observed state change to $\mathbf{x}^{j+1} \in \mathcal{X}'$, and where this behavior is indeed shown in a simulation. Hence, in contrast to the normally used stochastic simulation, we require that for states where at least two transitions are enabled, the decision between the alternatives is not taken randomly, but a specific transition is selected.

Thus, (standard) Petri nets have to be equipped with additional activation rules to force the switching of special transitions (to reach \mathbf{x}^{j+1} from \mathbf{x}^j), and to prevent all others from switching. For that, different types of additional activation rules are possible.

On the one hand, in [17, 21, 22] the concept of *priority relations* among the transitions of a network was introduced in order to allow the modeling of deterministic systems. Note that these priorities typically reflect

the rate of the corresponding reactions where the fastest reaction has highest priority and, thus, is taken. Priorities can prevent enabled transitions from switching: for each state, only a transition is *allowed to switch* if it is enabled and there is no other enabled transition with higher priority. In Marwan et al. [17] it is proposed to model such priorities with the help of partial orders \mathcal{O} on the transitions. We call $(\mathcal{P}, \mathcal{O})$ an *Petri net with priorities*, if $\mathcal{P} = (P, T, A, w)$ is a Petri net and \mathcal{O} a priority relation on $T \times T$.

On the other hand, the concept of control-arcs can be used to represent catalytic or inhibitory dependencies. An *extended Petri net* $\mathcal{P} = (P, T, (A \cup A_R \cup A_I), w)$ is a Petri net which has, besides the (standard) arcs in A , two additional sets of so-called control-arcs: the set of read-arcs $A_R \subset P \times T$ and the set of inhibitor-arcs $A_I \subset P \times T$. Also control-arcs can prevent enabled transitions from switching: an enabled transition $t \in T$ coupled with a read-arc (resp. an inhibitor-arc) to a place $p \in P$ can switch only if a token (resp. no token) is present in p . We denote the set of all arcs by $\mathcal{A} = A \cup A_R \cup A_I$.

For consistently integrating both concepts, priority relations and control-arcs, into the modeling framework, the difficulty is that both are concurrent concepts to force or prevent the switching of enabled transitions.

Let $(\mathcal{P}, \mathcal{O})$ be an *extended Petri net with priorities* with a network $\mathcal{P} = (P, T, (A \cup A_R \cup A_I), w)$ and a partial order $\mathcal{O} \subset T \times T$ on its transitions. To enforce a deterministic behavior, the network has to be equipped with appropriate control-arcs in $(A_R \cup A_I)$ and priorities in \mathcal{O} such that each state $\mathbf{x} \in \mathcal{X}$ has a unique successor $\text{succ}_{\mathcal{X}}(\mathbf{x})$, see [21] for more details. Extended Petri nets with priorities satisfying this property are said to be \mathcal{X} -deterministic. For our purpose, we consider a relaxed condition, namely that each experimentally observed state $\mathbf{x} \in \mathcal{X}'$ has a unique successor $\text{succ}_{\mathcal{X}'}(\mathbf{x}) \in \mathcal{X}'$, but we do not require this property for non-observed states $\mathbf{x} \in \mathcal{X} \setminus \mathcal{X}'$. To this end, in [23, 24], the notion of \mathcal{X}' -deterministic extended Petri nets is introduced as the desired output of an integrative reconstruction method, which show a prescribed behavior on the experimentally observed subset \mathcal{X}' of states: the reconstructed Petri nets (P, T, A, w) do not only contain enough transitions to reach the experimentally observed successors \mathbf{x}^{j+1} from \mathbf{x}^j , but exactly this transition will be selected (due to appropriate priorities and control-arcs) among all enabled ones in \mathbf{x}^j which is necessary to reach \mathbf{x}^{j+1} .

In Section 2, we survey results on the reconstruction of standard networks [8, 17], standard networks with priorities [17, 22], extended Petri nets [4, 5] and extended Petri nets with priorities [10, 11].

In Section 3, we discuss possible further extensions, future applications and the potential impact of the presented approaches.

2 Reconstruction approaches

We briefly describe the common input (P, \mathcal{X}') for network reconstruction, the main ideas, and the generated output of the different approaches to reconstruct Petri nets from experimental time-series data \mathcal{X}' .

Firstly, a set of components P (later represented by the set of places of the reconstructed networks) is chosen which is expected to be crucial for the studied phenomenon.

To perform an experiment, one first triggers the system in some state (by external stimuli like the change of nutrient concentrations or the exposition to some pathogens), to generate an initial state \mathbf{x}^1 . Then the system's response to the stimulation is observed and the resulting state changes are measured for all components at certain time points. This yields a sequence of (discrete or discretized¹) states $\mathbf{x}^j \in \mathbb{Z}^{|P|}$ reflecting the time-dependent response of the system to the stimulation in \mathbf{x}^1 , which typically terminates in a terminal state \mathbf{x}^k where no further changes are observed. The corresponding experiment is

$$\mathcal{X}'(\mathbf{x}^1, \mathbf{x}^k) = (\mathbf{x}^1, \dots, \mathbf{x}^k).$$

Several experiments starting from different initial states in a set $\mathcal{X}'_{ini} \subseteq \mathcal{X}'$, reporting the observed state changes for all components $p \in P$ at certain time points, and ending at different terminal states in a set $\mathcal{X}'_{term} \subseteq \mathcal{X}'$ describe the studied phenomenon, and yield experimental time-series data of the form

$$\mathcal{X}' = \{\mathcal{X}'(\mathbf{x}^1, \mathbf{x}^k) : \mathbf{x}^1 \in \mathcal{X}'_{ini}, \mathbf{x}^k \in \mathcal{X}'_{term}\}.$$

Thus, the input of the reconstruction approach is given by (P, \mathcal{X}') .

In the best case, two consecutively measured states $\mathbf{x}^j, \mathbf{x}^{j+1} \in \mathcal{X}'$ are also consecutive system states, i.e., \mathbf{x}^{j+1} is reached from \mathbf{x}^j by switching a single transition. This is, however, in general not the case (and depends on the chosen time points to measure the states in \mathcal{X}'), but \mathbf{x}^{j+1} may be obtained from \mathbf{x}^j by a switching sequence of some length, where the intermediate states are not reported in \mathcal{X}' .

For a successful reconstruction approach, the data \mathcal{X}' need to satisfy two properties:

- reproducibility (each $\mathbf{x}^j \in \mathcal{X}'$ has a unique observed successor state $\mathbf{x}^{j+1} \in \mathcal{X}'$) and
- monotonicity (the values of each place must not oscillate in the intermediate states between \mathbf{x}^j and \mathbf{x}^{j+1}).

¹Some measured data are of a discrete nature, e.g., a gene can be expressed or not, an enzyme can be present or not, a protein occurs in one of its conformational states. If some measured data are continuous, e.g., concentrations of certain metabolites, it is necessary to appropriately discretize the continuous data into finitely many discrete states. Hereby, the chosen level of discretization has to be fine enough to preserve all dynamic features of the time course (in particular all local maxima or minima of the values), but rough enough to be robust to noise in the experimental data.

Whereas reproducibility is obviously necessary, it was shown in [7] that monotonicity has to be required or, equivalently, that all essential responses are indeed reported in the experiments. Note that reproducibility can be tested in a preprocessing step, which includes strategies to modify the given experimental data in order to ensure the required properties, see [7, 22, 25] for details.

A network $\mathcal{P} = (P, T, A, w)$ fits the given data \mathcal{X}' when it is able to perform every observed state change from $\mathbf{x}^j \in \mathcal{X}'$ to $\mathbf{x}^{j+1} \in \mathcal{X}'$. More formally, with \mathcal{P} , an *incidence matrix* $C \in \mathbb{Z}^{|P| \times |T|}$ is associated whose rows correspond to places and whose columns $C_{\cdot t}$ to the *update vectors* \mathbf{r}^t of the transitions $t \in T$:

$$\mathbf{r}_p^t = C_{pt} := \begin{cases} -w(p, t) & \text{if } (p, t) \in A, \\ +w(t, p) & \text{if } (t, p) \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Reaching \mathbf{x}^{j+1} from \mathbf{x}^j by a switching sequence using the transitions from a subset $T' \subseteq T$ is equivalent to obtain the state vector \mathbf{x}^{j+1} from \mathbf{x}^j by adding the corresponding columns $C_{\cdot t}$ of C for all $t \in T'$:

$$\mathbf{x}^j + \sum_{t \in T'} C_{\cdot t} = \mathbf{x}^{j+1}.$$

Hence, T has to contain enough transitions to perform all experimentally observed switching sequences: \mathcal{P} is *conformal* with \mathcal{X}' if, for any two consecutive states $\mathbf{x}^j, \mathbf{x}^{j+1} \in \mathcal{X}'$, the linear equation system

$$\mathbf{x}^{j+1} - \mathbf{x}^j = C\lambda$$

has an integral solution $\lambda \in \mathbb{N}^{|T|}$ such that λ is the incidence vector of a sequence (t^1, \dots, t^m) of according transition switches.

Moreover, an extended Petri net is *catalytic conformal* with \mathcal{X}' and a Petri net with priorities is *\mathcal{X}' -deterministic* if there is a unique choice for the transitions to realize the switching sequences (due to appropriate control-arcs respectively appropriate control-arcs and priorities). Note that in all cases, the reconstructed networks have the same set P of places, as part of the input.

We next survey results on the reconstruction of

- standard networks [8, 17],
- standard networks with priorities [17, 22],
- extended Petri nets [4, 5] and
- extended Petri nets with priorities [10, 11].

2.1 Reconstructing standard networks

To reconstruct standard networks, the following approach is proposed in [8, 17].

Extraction of difference vectors. As initial step, extract the observed changes of states from the experimental data. For that, define the set

$$\mathcal{D} := \{\mathbf{d}^j = \mathbf{x}^{j+1} - \mathbf{x}^j : (\mathbf{x}^j, \mathbf{x}^{j+1}) \in \mathcal{X}'\}.$$

Decomposition of difference vectors. The first step is to describe the set of potential update vectors which might constitute the columns of C . Recall that two consecutively measured states $\mathbf{x}^j, \mathbf{x}^{j+1} \in \mathcal{X}'$ are not necessarily consecutive system states, but \mathbf{x}^{j+1} may be obtained from \mathbf{x}^j by some switching sequence whose intermediate states are not reported in \mathcal{X}' .

Due to monotonicity, the values of the elements cannot oscillate in the intermediate states between \mathbf{x}^j and \mathbf{x}^{j+1} . By [7], it suffices to decompose any $\mathbf{d}^j \in \mathcal{D}$ using only sign-compatible vectors from the following set

$$\text{Box}(\mathbf{d}^j) = \left\{ \mathbf{r} \in \mathbb{Z}^{|\mathcal{P}|} : \begin{array}{ll} 0 \leq r_p \leq d_p^j & \text{if } d_p^j > 0 \\ d_p^j \leq r_p \leq 0 & \text{if } d_p^j < 0 \\ r_p = 0 & \text{if } d_p^j = 0 \end{array} \right\} \setminus \{\mathbf{0}\}.$$

Due to [8], none of the reconstructed networks must contain a transition enabled at any of the observed terminal states \mathbf{x}^k of a time-series $(\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^k)$; hence all such vectors in $\text{Box}(\mathbf{d}^j)$ can be removed, resulting in a reduced set $\text{Box}_{red}(\mathbf{d}^j)$.

Next, we determine for any $\mathbf{d}^j \in \mathcal{D}$ the set $\Lambda(\mathbf{d}^j)$ of all integral solutions λ of the equation system

$$\mathbf{d}^j = \sum_{\mathbf{r}^t \in \text{Box}_{red}(\mathbf{d}^j)} \lambda_t \mathbf{r}^t, \quad \lambda_t \in \mathbb{Z}_+.$$

Due to [8], $\Lambda(\mathbf{d}^j)$ is non-empty if and only if \mathbf{d}^j is not enabled at any of the terminal states.

For each $\lambda \in \Lambda(\mathbf{d}^j)$, construct the (multi-)set $\mathcal{R}(\mathbf{d}^j, \lambda) = \{\mathbf{r}^t \in \text{Box}_{red}(\mathbf{d}^j) : \lambda_t \neq 0\}$ of update vectors used for this decomposition λ .

Composition of networks. Selecting exactly one solution $\lambda \in \Lambda(\mathbf{d}^j)$ for each $\mathbf{d}^j \in \mathcal{D}$ and taking the union of the corresponding sets $\mathcal{R}(\mathbf{d}^j, \lambda)$, we yield the columns $C_{\cdot t} = \mathbf{r}^t$ of an incidence matrix C of a (standard) network P . By construction, P is conformal with \mathcal{X}' since the selected update vectors (corresponding to the transitions of the network) are able to reproduce all experimentally observed state changes.

We obtain all such networks by taking all possible combinations of sets $\mathcal{R}(\mathbf{d}^j, \lambda)$ for different choices $\lambda \in \Lambda(\mathbf{d}^j)$ for all $\mathbf{d}^j \in \mathcal{D}$, see [17, 22].

2.2 Reconstructing standard networks with priorities

To generate standard networks with priorities, the first two steps of extracting and decomposing difference vectors are as in the case of standard networks without priorities. However, not all possible combinations of sets $\mathcal{R}(\mathbf{d}^j, \lambda)$ for different $\lambda \in \Lambda(\mathbf{d}^j)$ lead to \mathcal{X}' -deterministic networks. Here, it is necessary to consider all possible switching sequences for a decomposition $\lambda \in \Lambda(\mathbf{d}^j)$, to detect priority conflicts between different sequences, and to only compose selections of non-conflicting sequences to networks.

For that, the following approach is proposed in [17] (see [21, 22] for further details). We again extract difference vectors and decompose them as in the case of standard network reconstruction to determine the sets $\mathcal{R}(\mathbf{d}^j, \lambda)$ for all obtained difference vectors \mathbf{d}^j . Then, we proceed as follows:

Switching sequences and their conflicts. Every permutation $\pi = (\mathbf{r}^{t_1}, \dots, \mathbf{r}^{t_m})$ of the elements of a set $\mathcal{R}(\mathbf{d}^j, \lambda)$ gives rise to a sequence of intermediate states $\mathbf{x}^j = \mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^m, \mathbf{y}^{m+1} = \mathbf{x}^{j+1}$ with

$$\sigma = \sigma_{\pi, \lambda}(\mathbf{x}^j, \mathbf{d}^j) = ((\mathbf{y}^1, \mathbf{r}^{t_1}), (\mathbf{y}^2, \mathbf{r}^{t_2}), \dots, (\mathbf{y}^m, \mathbf{r}^{t_m}))$$

which induces a priority relation \mathcal{O}_σ since transition t_i resulting from \mathbf{r}^{t_i} is supposed to have highest priority in \mathbf{y}^i for $1 \leq i \leq m$. Two sequences σ and σ' are in *priority conflict* if there are update vectors $\mathbf{r}^t \neq \mathbf{r}^{t'}$ and intermediate states \mathbf{y}, \mathbf{y}' such that t, t' are enabled in \mathbf{y}, \mathbf{y}' and $(\mathbf{y}, \mathbf{r}^t) \in \sigma$ but $(\mathbf{y}', \mathbf{r}^{t'}) \in \sigma'$ (since this implies $t > t'$ in \mathcal{O}_σ but $t' > t$ in $\mathcal{O}_{\sigma'}$). Note that all sequences for the same $\mathbf{d}^j \in \mathcal{D}$ are mutually in conflict.

Selecting non-conflicting sequences. Construct a *priority conflict graph* \mathcal{G} whose nodes correspond to all possible sequences $\sigma_{\pi, \lambda}(\mathbf{x}^j, \mathbf{d}^j)$ and whose edges reflect their priority conflicts. In \mathcal{G} , all node subsets S are generated that select exactly one sequence $\sigma_{\pi, \lambda}(\mathbf{x}^j, \mathbf{d}^j)$ per difference vector $\mathbf{d}^j \in \mathcal{D}$ such that no priority conflicts occur between the selected sequences.

Remark. Each selected subset S corresponds to a so-called stable set in \mathcal{G} , all nodes corresponding to sequences for the same $\mathbf{d}^j \in \mathcal{D}$ to a so-called clique in \mathcal{G} . Note that we need to select one node from each such clique for a stable set S .

Composition of networks. Each subset S gives rise to a standard network $\mathcal{P} = (P, T, A, w)$, obtained by taking the union of the corresponding sets $\mathcal{R}(\mathbf{d}^j, \lambda)$ for each $\sigma = \sigma_{\pi, \lambda}(\mathbf{x}^j, \mathbf{d}^j) \in S$, which can be made \mathcal{X}' -deterministic by adding the priority relations \mathcal{O}_σ for all $\sigma \in S$, see [17, 22].

2.3 Reconstructing extended networks

To generate extended Petri nets without priorities, the decomposition of difference vectors has to be further adapted. As before, no transition of a returned network must be enabled at any of the observed terminal states \mathbf{x}^k of a time-series $(\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^k)$. However, this cannot be done by excluding potential update vectors, but instead by inserting control-arcs to disable the resulting transitions at the terminal states. This task can be done by using the procedure proposed in [4, 5].

Decomposition of difference vectors. Due to monotonicity, it suffices again to decompose any $\mathbf{d}^j \in \mathcal{D}$ using vectors from $\text{Box}(\mathbf{d}^j)$ only such that $\Lambda(\mathbf{d}^j)$ consists of all integral solutions λ of the equation system

$$\mathbf{d}^j = \sum_{\mathbf{r}^t \in \text{Box}(\mathbf{d}^j)} \lambda_t \mathbf{r}^t, \lambda_t \in \mathbb{Z}_+$$

and for each $\lambda \in \Lambda(\mathbf{d}^j)$, $\mathcal{R}(\mathbf{d}^j, \lambda)$ is constructed.

Composition of standard networks. We consider again permutations π of the elements of $\mathcal{R}(\mathbf{d}^j, \lambda)$ and the resulting switching sequences $\sigma_{\pi, \lambda}(\mathbf{x}^j, \mathbf{d}^j)$. Due to [5], such a sequence is catalytic conformal if and only if none of its intermediate states (including $\mathbf{x}^j = \mathbf{y}^1$) is an observed terminal state.

Selecting exactly one catalytic conformal sequence $\sigma_{\pi, \lambda}(\mathbf{x}^j, \mathbf{d}^j)$ for each $\mathbf{d}^j \in \mathcal{D}$ and taking the union of the corresponding sets $\mathcal{R}(\mathbf{d}^j, \lambda)$ yields the columns $C_t = \mathbf{r}^t$ of an incidence matrix C of a network. We obtain all such networks by taking all possible combinations.

Inserting control-arcs. It is assumed that control-arcs depend only on the presence or absence of a component, such that vectors $\mathbf{x} \in \mathcal{X}'$ are mapped to binary vectors in $\{0, 1\}^{|\mathcal{P}|}$ considering only the support of \mathbf{x} and thus can be tested in terms of propositional logic.

For this purpose, it is useful to encode the test by conjunctions of literals. Hereby, a literal can be any component x_p of a vector \mathbf{x} or its logical negation $\neg x_p$. A conjunction $\bigwedge_{x_p, \neg x_q \in P'} \text{literal}$ is true if and only if all positive variables $x_p \in P'$ have value greater than zero and all negated variables $\neg x_q \in P'$ have value zero. That way, each conjunction represents a pattern of components which need to be present or absent for the conjunction to be true. For details on how to find such a logical formula and the possible values to satisfy it, we refer to [5].

A conjunction can be modeled as a Petri net by connecting a transition to the places with read-arcs for every positive literal, and with inhibitor-arcs for every negative literal; the transition is enabled if and only if the conjunction is true.

2.4 Reconstructing extended networks with priorities

To generate extended Petri nets with priorities, the following approach is proposed in [10] and detailed in [11], based on previous works in [4, 5, 7, 8, 17].

Decomposition of difference vectors. Due to monotonicity, it suffices again to decompose any $\mathbf{d}^j \in \mathcal{D}$ using vectors from $\text{Box}(\mathbf{d}^j)$ only, to determine $\Lambda(\mathbf{d}^j)$ accordingly and to construct $\mathcal{R}(\mathbf{d}^j, \lambda)$ for each $\lambda \in \Lambda(\mathbf{d}^j)$.

Due to [10], $\text{Box}(\mathbf{d}^j)$ and $\Lambda(\mathbf{d}^j)$ are always non-empty since \mathbf{d}^j itself is always a solution due to the required reproducibility of the input data \mathcal{X}' (which particularly ensures $\mathbf{d}^j \neq \mathbf{0}$ for all $\mathbf{d}^j \in \mathcal{D}$, see also [25]).

Switching sequences and their conflicts. We consider again permutations π of the elements of $\mathcal{R}(\mathbf{d}^j, \lambda)$, the resulting switching sequences

$$\sigma = \sigma_{\pi, \lambda}(\mathbf{x}^j, \mathbf{d}^j) = ((\mathbf{y}^1, \mathbf{r}^{t_1}), (\mathbf{y}^2, \mathbf{r}^{t_2}), \dots, (\mathbf{y}^m, \mathbf{r}^{t_m}))$$

and the induced priority relations \mathcal{O}_σ . Recall that two sequences σ and σ' are in priority conflict if there are update vectors $\mathbf{r}^t \neq \mathbf{r}^{t'}$ and intermediate states \mathbf{y}, \mathbf{y}' such that t, t' are enabled in \mathbf{y}, \mathbf{y}' and $(\mathbf{y}, \mathbf{r}^t) \in \sigma$ but $(\mathbf{y}', \mathbf{r}^{t'}) \in \sigma'$. We have a weak (resp. strong) priority conflict if $\mathbf{y} \neq \mathbf{y}'$ (resp. $\mathbf{y} = \mathbf{y}'$) which can (resp. cannot) be resolved by adding appropriate control-arcs. Note that all sequences for the same $\mathbf{d}^j \in \mathcal{D}$ are in strong conflict.

Composition of standard networks. Construct a *priority conflict graph* \mathcal{G} whose nodes correspond to all catalytic conformal sequences $\sigma_{\pi, \lambda}(\mathbf{x}^j, \mathbf{d}^j)$ and whose edges reflect weak and strong priority conflicts. In \mathcal{G} , all node subsets S are generated that select exactly one sequence $\sigma_{\pi, \lambda}(\mathbf{x}^j, \mathbf{d}^j)$ per difference vector $\mathbf{d}^j \in \mathcal{D}$ such that no strong priority conflicts occur between the selected sequences. Each such subset S gives rise to a standard network $\mathcal{P} = (P, T, A, w)$ which is conformal with \mathcal{X}' and can be made \mathcal{X}' -deterministic.

Inserting control-arcs and adding priorities. To resolve a weak priority conflict between $\sigma, \sigma' \in S$ involving update vectors $\mathbf{r}^t \neq \mathbf{r}^{t'}$ and intermediate states $\mathbf{y} \neq \mathbf{y}'$, include

- either a read-arc $(p, t) \in A_R$ with weight $w(p, t) > y'_p$ for some p with $y_p > y'_p$ or
- an inhibitor-arc $(p, t) \in A_I$ with weight $w(p, t) < y_p$ for some p with $y_p < y'_p$

to disable transition t resulting from \mathbf{r}^t at \mathbf{y}' .

For each $\sigma \in S$, define \mathcal{O}_σ based on the so-constructed transition sequences to obtain the studied partial order to make the generated extended networks \mathcal{X}' -deterministic, see [10, 11] for details.

3 Discussion

Systems biology aims at the integrated experimental and theoretical analysis of cellular networks to obtain a holistic understanding of biological systems and processes. The large diversity of organisms and cell types and the huge variety of the involved biochemical processes requires advanced modeling and analysis approaches.

While metabolic networks are characterized by a flow of substance through biochemical intermediates that are interconverted into each other, regulatory or signal transduction networks function by a flow of information mediated by the reversible modification of functionally interacting components.

Since the information flow may occur through different molecular reaction mechanisms, it is not always trivial to unequivocally assign known biochemical reactions to a causal sequence of signaling events and vice versa. This is one reason why important control mechanisms like those for growth, differentiation, motility or apoptosis in eukaryotic cells are incompletely understood at the molecular systems level.

Reconstructing the functional interactions of genes and their products from experimental data by generating validated, composable, and predictive models of the underlying regulatory networks is one of the challenges of systems biology. To be useful for the reconstruction of causal interactions within regulatory networks from experimental data sets, it is suggested in [18] that a modeling framework should have the following properties. It should be able to

- represent the properties of the real system based on an unambiguous and strict syntax;
- reproduce qualitative and quantitative data in the particular formats as detected by different experimental techniques;
- connect functional elements or modules of arbitrary molecular complexity in the form of one single, coherent model;
- consistently represent the various molecular mechanisms or genetic interactions that occur in a living cell.

As exhibited in [13, 14, 18], all of these requirements are fulfilled by place/transition Petri nets and/or their various (stochastic or deterministic) extensions.

As mathematical structures, Petri nets allow the application of graph theoretical approaches and the mathematical treatment of formalized problems related to the molecular networks they are representing. This is particularly useful for the reconstruction and analysis of the regulatory networks studied in molecular cell biology and genetics.

So far, Petri nets as well as other modeling frameworks have mostly been used for bottom-up modeling approaches which are based on pre-defined molecular models of a given network: Experimental results are evaluated, and by heuristic reasoning, a network of biochemical reactions is established in the form of a classical biochemical pathway or in the form of an informal graphically displayed molecular interaction model. This molecular model is then translated into a Petri net to study the functional properties of the net, to run simulations, and to predict the behavior of the real network in response to perturbation.

Petri nets can also be obtained directly by algorithmically evaluating experimental data; such an approach is called network reconstruction, network inference, or reverse engineering. In this survey, we addressed the challenging problem of *Automatic Network Reconstruction*: reconstructing Petri net models from experimental time-series data by means of an exact, exclusively data-driven method. In the previous section, we gave an overview on approaches to reconstruct

- standard networks [7, 8, 17],
- standard networks with priorities [17, 22],
- extended Petri nets [4, 5], and
- extended Petri nets with priorities [10, 11].

In all cases, a *preprocessing* is in order to ensure the required properties of the input data.

On the one hand, we discussed the influence of the quality of the given experimental data in [7] showing that monotone data \mathcal{X}' are necessary (or, equivalently, that all essential responses are indeed reported in the experiments).

On the other hand, feasibility tests to test the given experimental data \mathcal{X}' for reproducibility have been developed (in [8] for standard networks and in [25] for extended Petri nets with priorities) to ensure that each observed state $\mathbf{x} \in \mathcal{X}'$ has a unique successor in \mathcal{X}' . If this test fails, the existence of previously undetected elements of regulatory function is proved since the given data do not suffice to explain the causal mechanisms behind the observed phenomenon [8, 25]. To handle infeasible cases, additional elements are added to P and the given experimental data \mathcal{X}' are extended by determining all possible values for the new elements in P such that any state $\mathbf{x} \in \mathcal{X}'$ with previously two different successors is split into different states each having a unique successor [8, 25]. Then, the regular reconstruction procedure is applied to every extended setting of P and \mathcal{X}' meeting these properties.

All here presented reconstruction approaches aim at generating *all* models of the studied type that fit with the given experimental data, to provide all possible alternatives of mechanisms behind the observed biological phenomena.

In the latter three cases, the generated networks are also able to exhibit the experimentally observed dynamic behavior in a simulation, since priorities and/or control-arcs are used as additional activation rules to force the switching of special transitions. In particular, the reconstruction of extended Petri nets with priorities [10] is an integrative method which generates models fully fitting the experimentally observed phenomena.

This typically causes a large number of solution alternatives, see e.g. [5, 10].

To keep this solution set as small as possible while still guaranteeing its completeness, a *postprocessing* should be applied to keep only "minimal" solutions in the sense that all other networks fitting the data contain the returned ones. For standard networks, minimality can be easily imposed by set inclusion of the transition sets [8, 17], for extended Petri nets with(out) priorities, an equivalence relation is defined on the underlying standard networks and minimal representatives are selected by means of priorities and/or control-arcs [23, 24].

Moreover, it would be desirable to exclude solution alternatives that contradict well-established biological knowledge or insight gained during previous rounds of the reconstruction process.

Although the here-surveyed reconstruction approaches are all exclusively data-driven methods, it could be advantageous to integrate some a priori biological knowledge already during the reconstruction process, instead of afterwards applying a biological verification of the reconstructed network alternatives as, e.g., done in [5].

We next discuss several possibilities proposed in [12] how certain biological knowledge could be integrated during the reconstruction process.

- *Indecomposability of difference vectors.* So far, all reconstruction approaches determine the set of potential update vectors which might constitute the networks by considering, for each $\mathbf{d}^j \in \mathcal{D}$, all sign-compatible vectors from the set $\text{Box}(\mathbf{d}^j)$.

In some cases, however, an observed response $\mathbf{d}^j \in \mathcal{D}$ might exactly correspond to a well-known biochemical reaction (including the correct stoichiometry) or to a well-known mechanism (that a certain trigger is detected by a suitable receptor).

In such cases, one could exclude the corresponding response $\mathbf{d}^j \in \mathcal{D}$ from decomposition and, instead, just define $\text{Box}(\mathbf{d}^j) := \{\mathbf{d}^j\}$ in accordance with the already established knowledge.

- *Benefiting from known P-invariants.* For a network $\mathcal{P} = (P, T, \mathcal{A}, w)$ with incidence matrix $C \in \mathbb{Z}^{|P| \times |T|}$, a P-invariant is an integer solution of the equation system

$$\mathbf{y}^T C = \mathbf{0}^T, \mathbf{y} \in \mathbb{N}^{|P|}, \mathbf{y} \neq \mathbf{0}.$$

P-invariants describe mass-preserving submodules: subsets $P' \subseteq P$ of places where the total number of tokens remains constant in each system state.

Typical examples of P-invariants are, e.g., different conformational states of a molecule, a receptor or a cell. If such P-invariants are known a priori within the experimentally studied set P of components, they can be provided as input in form of a set \mathcal{I}_P . This allows to reduce, for any $\mathbf{d}^j \in \mathcal{D}$, the set $\text{Box}(\mathbf{d}^j)$ of potential update vectors to decompose \mathbf{d}^j by including the restriction

$$\sum_{p \in P'} r_p = 0, \forall P' \in \mathcal{I}_P,$$

see also [8, 22] for details.

- *Treating equal difference vectors in the same way.* So far, all observed responses $\mathbf{d}^j \in \mathcal{D}$ are decomposed independently from each other. To generate standard networks, exactly one solution $\lambda \in \Lambda(\mathbf{d}^j)$ for each $\mathbf{d}^j \in \mathcal{D}$ is selected, for networks with priorities, permutations π of the elements of the corresponding sets $\mathcal{R}(\mathbf{d}^j, \lambda)$ give rise to sequences $\sigma_{\pi, \lambda}(\mathbf{x}^j, \mathbf{d}^j)$, and non-conflicting sequences are selected.

If two difference vectors $\mathbf{d}^i, \mathbf{d}^j \in \mathcal{D}$ are equal, we have $\text{Box}(\mathbf{d}^i) = \text{Box}(\mathbf{d}^j)$ and it would be natural to require that both $\mathbf{d}^i, \mathbf{d}^j$ are decomposed in the same way (i.e., by the same $\lambda \in \Lambda(\mathbf{d}^i) = \Lambda(\mathbf{d}^j)$) and that the involved reactions are applied in the same order (i.e., by the same permutation π of the elements of the corresponding sets $\mathcal{R}(\mathbf{d}^i, \lambda) = \mathcal{R}(\mathbf{d}^j, \lambda)$).

The latter could be achieved by adding (strong) priority conflicts between all other sequences stemming from a pair $\mathbf{d}^i, \mathbf{d}^j$ of equal difference vectors while creating the priority conflict graph, since no two sequences in (strong) priority conflict are selected for the same network.

- *Knowledge on reaction rates, catalysts or inhibitors.* In the case of reconstructing extended Petri nets with priorities, sequences σ, σ' in weak priority conflict might be selected for a network.

To resolve a weak priority conflict between σ, σ' involving update vectors $\mathbf{r}^t \neq \mathbf{r}^{t'}$ and intermediate states $\mathbf{y} \neq \mathbf{y}'$, either transition t has to be disabled at \mathbf{y}' or transition t' at \mathbf{y} , while the decision between t and t' on the other state can be handled by a priority.

Here, pre-knowledge about the reaction rates of t and t' (e.g. gained from the time-scales during the experiments) could help to decide whether $t > t'$ or $t < t'$ better reflects the reality.

Furthermore, to resolve the conflict on the remaining state, at least one control-arc of the following form is included:

- a read-arc (p, t) with weight $w(p, t) > y'_p$ for some p with $y_p > y'_p$ or
- an inhibitor-arc (p, t) with weight $w(p, t) < y_p$ for some p with $y_p < y'_p$.

A priori knowledge on catalysts or inhibitors of the studied reactions $r^t, r^{t'}$ can help to reduce the potential alternatives of control-arcs and, thus, the total number of generated network alternatives.

Note that a priori knowledge can also be obtained in terms of structural insights of an interplay of components gained in previous reconstruction steps. In fact, it is possible to step-wise refine Petri net models since places and transitions may summarize (lump) subnetworks of biochemical reactions of known or unknown molecular detail, see [18] for further details.

So, it is possible to run a reconstruction process on a currently known input and, as soon as it turns out that further components are crucial for the studied phenomenon or new experimental data becomes available, it is possible to re-run the reconstruction procedure on an extended input setting (P, \mathcal{X}') , thereby preserving insights and approved subnetworks gained in the previous reconstruction round(s).

Impact of Automatic Network Reconstruction.

To summarize, we surveyed an exact combinatorial approach to generate a complete list of all Petri nets that are conformal with the given data [8, 17] or even display the experimentally observed behavior in a simulation [4, 5, 10, 11].

For standard and extended networks, we verified the correctness of the algorithm with the help of some small biological instances [5, 6]. For standard networks, we provided an implementation of the reconstruction approach using Answer Set Programming [3].

The further goal is to provide such an implementation also for extended Petri nets with and without priorities, hereby including both options of integrating previous knowledge during the reconstruction as proposed in [12] or of generating the whole set of all technically possible solution alternatives.

Finally, we plan to apply the presented reconstruction approaches to different biological experimental data, e.g., to

- cell differentiation processes depending on the correlated expression of developmentally regulated genes in *Physarum polycephalum* (a model organism for cell differentiation) like in [9];
- sulfur metabolism in wheat and its implications to the wheat grain protein composition like described in [20].

We indeed expect an important impact of Automatic Network Reconstruction in order to support the integrated experimental and theoretical analysis of biological systems and processes towards their holistic understanding.

Moreover, note that the here surveyed reconstruction procedures have a large potential as they can be applied in various contexts:

- *Reverse engineering:* The main application consists in determining all molecular or regulatory networks or networks of causal dependencies fitting given biological experimental data.

Note that this includes to detect causal interactions between host and pathogen or between cells of a tissue or tissues of an organ, with or without drug application.

- *Experimental design:* Once the reverse engineering step is performed, it is possible to analyze the obtained solution set in order to define further experiments to rule out inappropriate solution alternatives.

This can be done by defining experiments that allow and are sufficient to unequivocally distinguish between alternatives of a complete list of alternative networks, thereby avoiding to perform unnecessary or to miss necessary experiments.

- *Step-wise refining models:* In an iterative process of experimentation and computation, it is possible to obtain a hierarchically defined Petri net where places and transitions are refined into subnetworks, resolving individual biochemical reactions at any level of kinetic detail while other nodes of the same net may still represent more complex processes like the activation of a cluster of differentially expressed genes or the assumption of a physiological state of the entire cell.

This is possible since Petri nets provide a powerful modeling tool: Even if many molecular details are still unknown, models can be obtained and simulations run to predict or verify experimental results.

- *Systems design:* In synthetic biology, the goal is to design systems with predefined structural motives or a prescribed dynamic behavior.

This can be achieved by creating the complete list of all alternative networks, or changes in the structure of a given network, by applying reconstruction procedures to time-series data that describe those predefined motives or exhibit the prescribed dynamic behavior.

References

- [1] Andrec, M., Kholodenko, B., Levy, R., Sontag, E.: Interference of signaling and gene regulatory networks by steady-state perturbation experiments: structure and accuracy. *Journal of Theoretical Biology* **232**, 427–441 (2005)
- [2] M. Chen and W. Hofestädt, Quantitative Petri net model fo gene regulated metabolic networks in the cell, *In Silico Biology* **3**, 347–365 (2003)
- [3] M. Durzinsky, W. Marwan, M. Ostrowski, T. Schaub, A. Wagler. Automatic network reconstruction using ASP. *Th. and Practice of Logic Progr.* 11:749–766, 2011.
- [4] M. Durzinsky, W. Marwan, A. Wagler. Reconstructing extended Petri nets from time series data and its application to signal transduction and to gene regulatory networks. *BMC Systems Biology* 5, 2011.
- [5] M. Durzinsky, W. Marwan, A. Wagler. Reconstruction of extended Petri nets from time-series data by using logical control functions. *J. of Math. Biology* 66:203–223, 2013.
- [6] M. Durzinsky, W. Marwan, A. Wagler, R. Weismantel. Automatic reconstruction of molecular and genetic networks from experimental time series data. *BioSystems* 93:181-190, 2008.
- [7] M. Durzinsky, A. Wagler, R. Weismantel. A combinatorial approach to reconstruct Petri nets from experimental data. *Lect. Notes in Comp. Science* 5307:328–346, 2008 (Special Issue CMSB 2008).
- [8] M. Durzinsky, A. Wagler, R. Weismantel. An algorithmic framework for network reconstruction. *J. of Theor. Comp. Science* 412:2800–2815, 2011.
- [9] B. Ebeling, X.-K. Hoffmann, V. Rätzel, J. Tesmer, W. Marwan. Physarum polycephalum mutants in the photocontrol of sporulation display altered patterns in the correlated expression of developmentally regulated genes. *Develop. Growth Differ.* 55 (2013) 247–259
- [10] M. Favre, A. Wagler. Reconstructing \mathcal{X}' -deterministic extended Petri nets from experimental time-series data \mathcal{X}' (extended abstract). *CEUR Workshop Proceedings* 988 (2013) 45-59 (Special Issue BioPPN 2013).
- [11] M. Favre, A. Wagler. Reconstructing extended Petri nets with priorities: Handling priority conflicts revisited. Submitted to *Lect. Notes in Comp. Science* (Special Issue ISCO 2014).
- [12] M. Favre, W. Marwan, A. Wagler. Integrating a priori biological knowledge in Automatic Network Reconstruction. Manuscript
- [13] I. Koch, M. Heiner. Petri nets. In: B.H. Junker, F. Schreiber (eds.), *Biological Network Analysis*, Wiley Book Series on Bioinformatics, 139–179, 2007.
- [14] I. Koch, F. Schreiber, W. Reisig (eds.). *Modeling in Systems Biology*, Computational Biology 16, Springer London, 2010.
- [15] Krishna, R., Guo, S., A partial granger causality approach to explore causal networks derived from multi-parameter data, *Computational Methods in Systems Biology* **5307**, 9–27 (2008)
- [16] Laubenbacher, R., Stigler, B.: A computational algebra approach to reverse engineering of gene regulatory networks. *Journal of Theoretical Biology* **229**, 523–537 (2005)
- [17] W. Marwan, A. Wagler, R. Weismantel. A mathematical approach to solve the network reconstruction problem. *Math. Methods of Oper. Research* 67:117–132, 2008.
- [18] W. Marwan, A. Wagler, R. Weismantel. Petri nets as a framework for the reconstruction and analysis of signal transduction pathways and regulatory networks, *Natural Computing* 10:639–654, 2011
- [19] J.W. Pinney, R.D. Westhead, G.A. McConkey. Petri net representations in systems biology. *Biochem. Soc. Trans.* 31:1513–1515, 2003.
- [20] Ravel C, Plessis A, Duchateau N, Dardevet M, Bordes J, Rhazi L, Balfourier F, Martre P. Genetic mapping of wheat grain protein composition reveals that gliadin and glutenin composition are trans-regulated by different chromosomic regions. *Journal of Experimental Botany*. In Press.
- [21] L.M. Torres, A. Wagler. Encoding the dynamics of deterministic systems. *Math. Methods of Oper. Research* 73:281–300, 2011.
- [22] A. Wagler. Prediction of network structure, In: *Modeling in Systems Biology*, I. Koch, F. Schreiber, W. Reisig (eds.) Computational Biology 16, Springer London, 309–338, 2010,
- [23] A. Wagler, J.-T. Wegener. On minimality and equivalence of Petri nets (extended abstract). *CEUR Workshop Proceedings* 928 (2012) 382-393 (Special Issue CS&P 2012)
- [24] A. Wagler, J.-T. Wegener. On minimality and equivalence of Petri nets. *Fundamenta Informaticae*, DOI 10.3233/FI-2012-922.
- [25] A. Wagler, J.-T. Wegener. Preprocessing for network reconstruction: Feasibility test and handling infeasibility (extended abstract). *CEUR Workshop Proceedings* 1032 (2013) 434-447 (Special Issue CS&P 2013)