



**HAL**  
open science

# **Inference of ecological networks and possibilistic dynamics based on Boolean networks from observations and prior knowledge**

Loïc Paulevé, Cédric Gaucherel

► **To cite this version:**

Loïc Paulevé, Cédric Gaucherel. Inference of ecological networks and possibilistic dynamics based on Boolean networks from observations and prior knowledge. *Methods in Ecology and Evolution*, 2025, 16 (8), pp.1851-1867. <10.1111/2041-210x.70090>. <hal-05143187>

**HAL Id: hal-05143187**

**<https://hal.science/hal-05143187v1>**

Submitted on 3 Jul 2025

**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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

## RESEARCH ARTICLE

# Inference of ecological networks and possibilistic dynamics based on Boolean networks from observations and prior knowledge

 Loïc Paulevé<sup>1</sup>  | Cédric Gaucherel<sup>2</sup> 
<sup>1</sup>University of Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, Talence, France

<sup>2</sup>AMAP—INRAE, CIRAD, CNRS, IRD, Montpellier University, Montpellier, France
**Correspondence**

Loïc Paulevé

Email: [loic.pauleve@labri.fr](mailto:loic.pauleve@labri.fr)**Funding information**

Agence Nationale de la Recherche, Grant/Award Number: ANR-20-CE45-0001 and ANR-23-CE45-0008

**Handling Editor:** Robin Boyd**Abstract**

1. Being able to infer the interactions between a set of species from observations of the system is of paramount importance to obtain explanatory and predictive models in ecology. We tackled this challenge by employing qualitative modelling frameworks and logic methods for the synthesis of mathematical models that can integrate both observations and expert knowledge on the system.
2. Boolean networks is a qualitative modelling framework, which enables reasoning exhaustively on possible dynamics of the system. After devising a formal link between ecological networks and the causal structure of Boolean networks, we applied a generic model synthesis engine to infer Boolean models that are able to reproduce the observed dynamics of a protist community and of a planktonic ecosystem.
3. Our inference method supports optimization criteria to derive the most parsimonious and most precise models. It is also able to integrate prior knowledge on the ecological network, adding constraints on impossible interactions, which is necessary to obtain realistic predictions. Such constraints may, however, prove to be too strict, in which case our method is able to conclude on the absence of a model compatible with both the observations and the input hypotheses.
4. We demonstrated our methodology on experimental data of a protist community and of a planktonic ecosystem and showed in each case its ability to recover essential and sufficient ecological interactions to explain the observed dynamics.

**KEYWORDS**

community ecology, food webs, modelling, species interactions

**1 | INTRODUCTION**

Understanding and predicting how natural communities are formed, often represented as species networks, is critical in ecology. Such

communities have been studied by sequentially adding species or by invading an already formed community by a group of pre-selected species (Serván & Allesina, 2021). Such structural changes are highly conditioning the resulting interaction networks (Fukami, 2015).

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

However, it is often highly difficult to guess the interaction network from field observations, except in rare laboratory experiments, thus suggesting developing inference methods aiming at identifying the interaction network leading to specific observed interactions (Milns et al., 2010; Pocock et al., 2021). Such networks are mandatory to anticipate the community dynamics (Gaucherel, Fayolle, et al., 2024; Veloz, 2019).

Although many models of species community have been developed (Baldan et al., 2018; Serván & Allesina, 2021), inference methods have been much less developed. Inference frameworks mainly build on Bayesian networks aiming at identifying the interactions populating such networks Milns et al. (2010) or the interaction strength associated with a community (Pocock et al., 2021). They can even integrate expert knowledge and use supervised learning machines to improve network inference (Auclair et al., 2017). Usually, such existing methods require a large amount of observations and knowledge on the system studied, to infer the interaction network and to calibrate the model. Whenever accounting for species (population) abundances and/or the interaction intensities (fluxes), statistical inference methods often use quantitative data for their parameterization.

Searching for qualitative methods would be an original step in the way of identifying ecological networks (Dambacher et al., 2003; Gaucherel, Fayolle, et al., 2024). In such a case, the qualitative network of interactions (and interaction signs) would be inferred before going into the details of the strength or probabilities of the interactions, requiring more data. Several studies show that qualitative modelling frameworks, such as Boolean networks, greatly facilitate the inference of models of biological systems (Barroso-Bergadá et al., 2020; Yu et al., 2004), and can even enable reasoning over possible quantitative realization, without having to infer parameters (Gaucherel & Pommereau, 2019; Paulevé et al., 2020).

In this paper, we demonstrate the feasibility of inferring ecological networks through the automatic design of Boolean dynamical models by integrating both observations of (community) system transitions and prior knowledge. To that aim, we first established connections between Boolean network (BN) theory and trophic networks and show how global features such as the acyclicity and (in) sufficiency of trophic interactions can be translated into constraints over the architecture of a so-called influence graph of BNs (similar to the ecological interaction network). This enabled tailoring the generic method for Boolean network synthesis implemented in the BoNESis software (Chevalier, Boyenval, et al., 2024), which relies on logic programming to automatically deduce Boolean models compatible with prior constraints on their architecture and possess specified dynamical properties, including reproducing given transitions, trajectories, and steady states.

An important facet of BNs is their ability to access the exhaustive possible dynamics of the studied systems. Indeed, because the number of configurations of a BN is finite, one can explore all possible trajectories allowed by it, and notably, reason on the stable states, or attractors, of the system. This way of computing all possible pathways of a system dynamics is called *possibilism*, providing a rich

alternative to the deterministic and probabilistic models (Gaucherel, Cosme, et al., 2024).

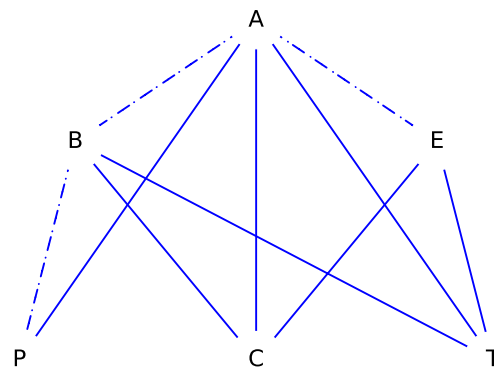
We applied our methodology to the inference of a protist community network based on the observations gathered from one of the most exhaustive experiments ever in ecology (Weatherby et al., 1998) and have been well-studied since then (de Goër de Herve et al., 2022; Werner & Peacor, 2003). Our inference strategy requires first to build a qualitative specification of dynamics and specify constraints on the expected trophic networks (e.g. with a vertical hierarchy), related to expert knowledge. Then, employing BoNESis, one can deduce compatible Boolean networks, and then the underlying trophic networks. We explored different optimization criteria, including identifying Boolean networks that reproduce the dynamics as closely as possible to the observations. In that sense, the inferred trophic networks are justified by the existence of a qualitative dynamical model that is able to replicate the observed transitions (Gaucherel & Pommereau, 2019).

In order to emphasize the flexibility of the framework to integrate prior expert knowledge on the ecological system, we performed the inference with (1) a zero knowledge inference, only requiring acyclic trophic networks, (2) prior on trophic networks from Weatherby et al. (1998) and (3) extended prior on trophic networks from de Goër de Herve et al. (2022). This resulted in inferred community networks from increasing fits with the observed state transitions and ecological knowledge. Models are required to have a perfect fit on observations (all system transitions are retrieved) and have minimal deviations, that is, additional non-observed transitions. We further demonstrated that the inference method can be applied to a different and an even more realistic setting: the inference of a planktonic ecosystem from observed time series. In this case study, we emphasized the ability of our framework to express different dynamical properties, and its relevance whenever we have few observations on the system, leading to numerous candidate models.

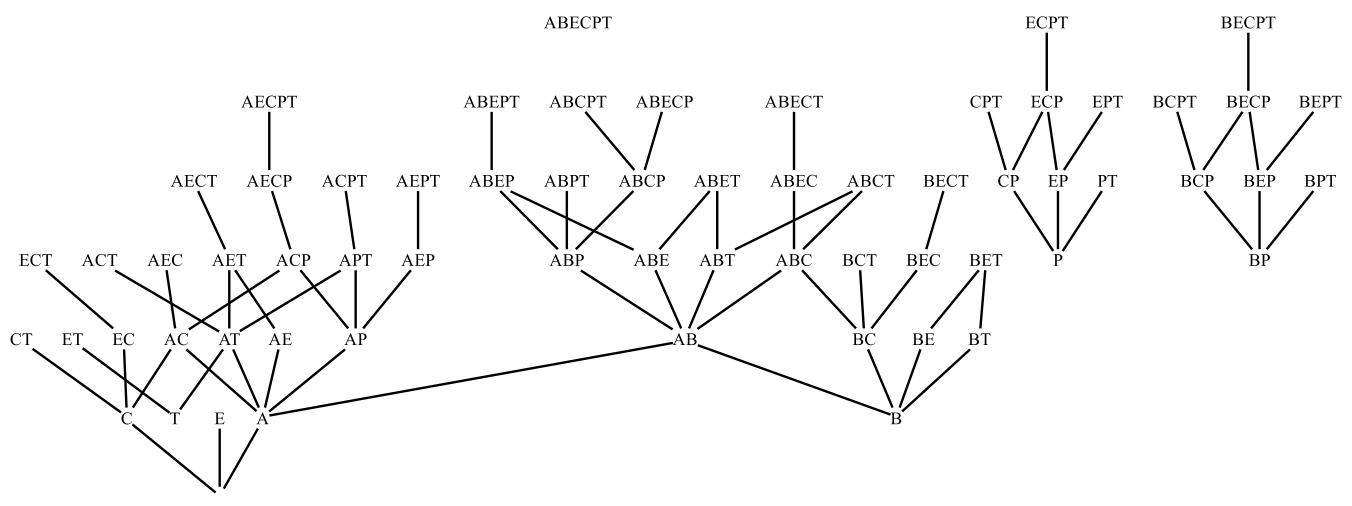
## 2 | MATERIALS AND METHODS

### 2.1 | The protist community

As a case study, we focussed on the seminal protist experiment published by the University of Sheffield (Law et al., 2000; Warren et al., 2003, 2006; Weatherby et al., 1998). The experiment consisted of registering the successive states and transitions of a protist community consisting of up to six species, chosen to build a trophic network involving/ various strategies and sizes. The selected species were *Amoeba proteus* (noted A; ~200 µm), *Blepharisma japonicum* (B; ~200 µm), *Paramecium caudatum* (P; ~150 µm), *Euplotes patella* (E; ~140 µm), *Colpidium striatum* (C; ~70 µm) and *Tetrahymena pyriformis* (T; ~35 µm), of the largest to the smallest sizes, feeding on unlimited bacterial resources. These species have been growing in a controlled (laboratory) environment and can only survive or die without invasion. Their trophic and possible competition interactions, represented in Figure 1a, were first based on expert knowledge about their feeding



(a) WLW prior trophic network



(b) Transitions observed experimentally

**FIGURE 1** (a) WLW prior trophic network (Law et al., 2000) between six protist species feeding on unlimited bacteria resource (not represented). Dashed-dotted edges indicate insufficient trophic interactions. (b) Qualitative transitions derived from experimental observations. Chronology and reading direction are from top to bottom.

habits, previously tested, and then on confirmed transitions observed between pools of species (Warren et al., 2006; Weatherby et al. 1998).

Each of the 63 combinations of protist species has been tested and replicated six times, with previously stabilized stock cultures and about a hundred individuals of each species. The resulting 378 microcosms were then sampled on Day 17 and every 28 days until some chosen stopping criteria were reached. The observations stopped when either (i) all species went extinct, or (ii) only one species not able to persist alone on bacteria, (iii) only one species able to persist remained or (iv) the species composition was unchanged for three successive samples. Most microcosms (93%) met one of these stopping criteria by 101 days in the original experiment (Weatherby et al., 1998). Here, we based our analysis on both the expert knowledge trophic network (Warren et al., 2003, 2006; Weatherby et al., 1998) and the full time-series information available to reconstruct all qualitative time series and pathways of the protist community (de Goër de Herve et al., 2022; Warren

et al., 2003). Indeed, we focussed on presence/absence data of each species, previously confirmed by systematic scanning of the entire microcosm states after 20–30 times magnification, to draw a new trophic network with competition events in agreement with the observed transitions. The experiment revealed highly non-consistent community pathways and no systematic final states or timing for transitions (Warren et al., 2003; Weatherby et al., 1998). Most of the time, the observed transitions were not reproducible, which appears quite usual but is rigorously confirmed in this experiment. At least 32 of the 63 experiments displayed multiple and different final compositions and pathways. Observed 71 transitions are summarized in Figure 1b. Each transition modifies the state of only one species. Furthermore, states with ingoing transitions but without outgoing observed transitions represent *steady states* of the system. They correspond to species B, P, T alone, BP, and total extinction ( $\_$  state). No observed transitions from the state where all species are present: indeed, this state is not lasting,

as many prey species are likely to rapidly disappear when all predator species are simultaneously present.

## 2.2 | Time series of planktonic ecosystem

The data on the planktonic ecosystem come from marshes of the French Atlantic coast (Charente-Maritime region), which are freshwater ecosystems from unreplenished drained marshes. Samples of the planktonic trophic network are recorded at two stations (stations A and B) on successive weeks during winter and spring 2012 to reconstruct the network trajectories and their environment over time (Masclaux et al., 2014). The planktonic trophic network is composed of seven species (sorted by sizes) and is completed by abiotic and anthropogenic components characterizing the environment such as nitrates and Dissolved Organic Matter. The trophic network regimes change during the winter to spring transition, from biological winters to ecosystems qualified as multivorous networks. To these observations, we added an expected theoretical pathway based on expert elicitation, basically following the previously mentioned winter to spring transition. The trophic regimes identified here are theoretically shifting from biological winters, followed by herbivorous networks, to finally reach complex multivorous networks with three levels of multivory (weakly multivorous, multivorous and highly multivorous) (Masclaux et al., 2014).

The data reflect transient and stable regimes depending on environmental changes (and internal influences among species), which we modelled here with the presence or absence of bacteria and nitrates on which feed the basal species. The time series are summarized in Figure 2.

condition	Station A			Station B				Theoretical					
	A1	A2	A3	B1	B2	B3	B4	T1	T2	T3	T4	T5	T6
<b>Bact</b>	0	1	1	0	1	1	1	0	0	0	1	1	1
<b>Nit</b>	1	1	0	1	1	1	0	1	1	1	0	0	0
<b>PicoP</b>	0	1	1	0	1	1	1	0	0	0	1	1	1
<b>NanoP</b>	0	0	0	0	0	0	1	0	1	1	1	0	0
<b>MicrP</b>	0	1	1	0	1	1	0	0	1	1	0	0	0
<b>Proto</b>	0	0	1	0	1	0	0	0	0	0	1	1	0
<b>MicrZ</b>	0	1	0	0	0	1	1	0	0	1	1	1	0
<b>MesoZ</b>	0	1	1	0	1	1	1	0	0	1	1	0	0

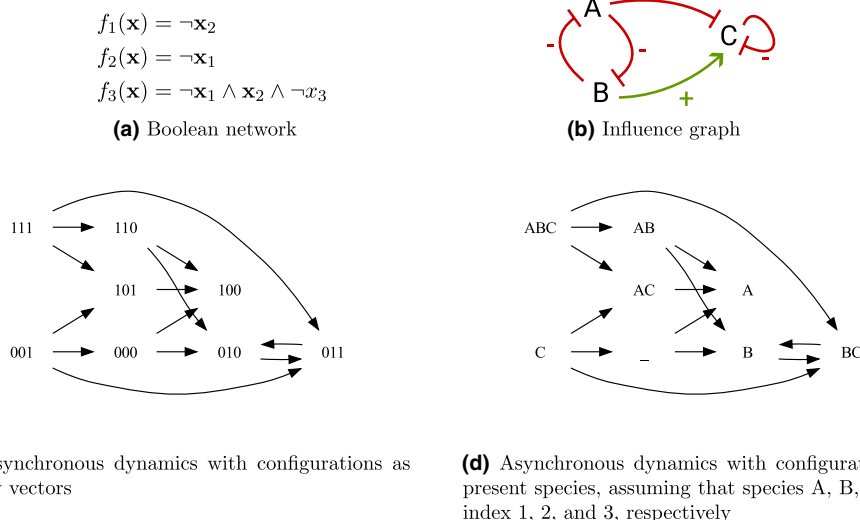
**FIGURE 2** Observed (at stations A and B) and theoretical time series of the planktonic ecosystem used for model inference. The three pathways are listed as matrices of ecosystem components (rows) versus the ecosystem regimes observed (columns) in the three pathways.

## 2.3 | Boolean networks and influence graphs

*Boolean networks* (BNs) are discrete dynamical models that describe the change in activity state (on/off) of species using Boolean functions (Kauffman, 1969; Thomas, 1973). Formally, a BN  $f$  among  $n$  species is defined by  $n$  Boolean functions  $f_i: \mathbb{B}^n \rightarrow \mathbb{B}$  for  $i = 1, \dots, n$ . Each function specifies the target Boolean state of the related species according to the Boolean state of the other species in the network. The state of all the species is represented by a *configuration* which is a Boolean vector of dimension  $n$ , i.e., within  $\mathbb{B}^n$ .

The functions of a BN are typically represented using propositional logic ( $\neg$  is negation,  $\wedge$  conjunction, and  $\vee$  disjunction). An example is given with its influence graph in Figure 3a.

BNs are typically employed to predict successions of configurations and transitions, generating *trajectories*. Trajectories are computed from  $f$  by following an *update mode*. For example, with the *synchronous mode*, all species are updated in parallel, leading to a unique transition from any configuration  $\mathbf{x} \in \mathbb{B}^n$  to the configuration  $f_1(\mathbf{x})f_2(\mathbf{x}) \dots f_n(\mathbf{x})$ . In the scope of this paper, we consider trajectories of BNs computed using the (fully) *asynchronous mode*, where only one species changes of state at a time. In this case, in general, several outgoing transitions are possible from a single configuration, leading to non-deterministic or undeterministic dynamics, sometimes called possibilistic (i.e. non-deterministic yet systematic). The non-determinism typically encompasses probabilities and stochasticity of transitions, as well as uncertainty on duration of state changes (Paulevé et al., 2020). BN dynamics can be captured by a *transition graph*, which is a directed graph  $\text{DynA}(f)$  between configurations  $\mathbb{B}^n$  such that there is an edge from a configuration  $x$  to  $y$  if  $y$  is the result of the update of a single species  $i$  according to  $f_i$ :



**FIGURE 3** Example of Boolean network with corresponding influence graph and asynchronous dynamics.

$\forall x, y \in \mathbb{B}^n, x \rightarrow y \in \text{DynA}(f) \Leftrightarrow \exists i \in [n], \Delta(x, y) = \{i\}$  and  $y_i = f_i(x)$ , where  $\Delta(x, y) = \{i \in [n] | x_i \neq y_i\}$  is the set of species whose states differ between configurations  $x$  and  $y$ .

Figure 3c gives an example of a transition graph, with configurations represented as binary vectors. In this paper, we also represent configurations by the species that are in State 1, as shown in Figure 3d.

An *attractor* is a non-empty set of configurations  $A \subseteq \mathbb{B}^n$  that no trajectory can escape, and such that any pair of configurations are connected by a trajectory. Attractors can be composed of a single configuration, called *fixed points*, where the state of each species cannot change; or of multiple configurations, leading to sustained cyclic behaviours. They correspond to the *terminal strongly connected components* of the graph  $\text{DynA}(f)$  and model the long-term behaviours of the system.

The asynchronous dynamics of the BN of Figure 3a has two attractors: the fixed point  $\{100\}$ , and the cyclic attractor  $\{010, 011\}$ .

### 2.3.1 | Influence graph

In general, the function  $f_i$  of a species  $i$  does not depend on the state of all the species of the network. The *influence graph* of  $f$  captures the (signed) dependencies between the functions of the species: there is positive (resp. negative) influence of  $j$  to  $i$  if there exists at least one configuration  $x$  in which the sole increase of  $j$  change the value of  $f_i$  from 0 to 1 (resp. 1 to 0):

$$j \xrightarrow{+} i \in G(f) \Leftrightarrow \exists x \in \mathbb{B}^n, f_i(x_1 \dots x_{j-1} 0 x_{j+1} \dots x_n) < f_i(x_1 \dots x_{j-1} 1 x_{j+1} \dots x_n),$$

$$j \xrightarrow{-} i \in G(f) \Leftrightarrow \exists x \in \mathbb{B}^n, f_i(x_1 \dots x_{j-1} 0 x_{j+1} \dots x_n) > f_i(x_1 \dots x_{j-1} 1 x_{j+1} \dots x_n).$$

A BN  $f$  such that there exists  $i, j$  with both  $j \xrightarrow{+} i \in G(f)$  and  $j \xrightarrow{-} i \in G(f)$  is *non-monotone*, otherwise it is *locally monotone*. In the scope of this paper, we do not consider non-monotone BNs and assume that

for any species  $j$  it is either an activator ( $j \xrightarrow{+} i \in G(f)$ ), or an inhibitor ( $j \xrightarrow{-} i \in G(f)$ ) of another species  $i$ , or  $j$  has no direct influence on  $i$ .

An example of an influence graph is given by Figure 3b. Positive edges are drawn in green, and negative edges are drawn in red with a bar arrow. The example BN is locally monotone.

### 2.3.2 | Note on figures

To help distinguish between the three types of graphs we use, ecological (and trophic) networks are drawn with blue edges, influence graphs are drawn using green and red edges, and transition graphs are drawn with black edges. Note that in several figures, additional colours are used to highlight specific edges. All these graphs are directed; however, for the sake of readability, we omit arrow tips when direction is from top to bottom.

## 2.4 | Inference of Boolean networks from structure and dynamics

We focus on the inverse problem of identifying BNs whose influence graph is included in a prior signed digraph and whose asynchronous dynamics include the observed transitions and fixed points. More formally, the input of the inference problem is a signed digraph  $\mathcal{G}$ , a set of asynchronous transitions  $T \subseteq \{x \rightarrow y | x, y \in \mathbb{B}^n, |\Delta(x, y)| = 1\}$ , and a set of configurations  $S \subseteq \mathbb{B}^n$  that must be fixed points. The inference consists in finding BNs  $f$  such that  $G(f) \subseteq \mathcal{G}$ ,  $T \subseteq \text{DynA}(f)$ , and for each  $x \in S$ ,  $f(x) = x$ . Ecologists and modellers often use the Akaike or Bayesian information criteria to estimate the quality of a statistical model and its associated prediction error for a specific dataset (Burnham & Anderson, 2002). However, such indices are not appropriate for our piece of work, in particular because they concern quantitative data and the absence of parameters in our qualitative models.

Many different BNs may fulfil these properties. We considered two measures:

- *Sparsity*: we will focus on BNs requiring a minimum set of influences (interactions) between species. The minimalist criterion is computed with respect to set inclusions: a set of influences is minimal if there is no BN using a strict subset of them while verifying the required dynamical properties. Note that there may exist several minimal BNs.
- *Deviation*: we are interested in having models reproducing the observations as closely as possible. While we impose that a BN has *at least* the observed transitions, we can compute how many additional *unobserved* transitions are predicted by the model from the observed states. The less deviation, the better.

In this work, we employed the tool BoNESIS (Chevalier et al., 2019; Chevalier, Boyenval, et al., 2024), based on logic Answer-Set Programming (ASP) and CLINGO solver (Baral, 2003; Gebser et al., 2014). The tool is restricted to the inference of locally monotone BNs, but can easily be extended to the above-mentioned dynamical constraints and optimization criteria.

## 2.5 | Methodology

For the protists case study, our general network inference strategy relies on the identification of BNs that match with observations and optional constraints on the structure of influences, and that are optimal with respect to sparsity and deviation. Then, from the identified BNs, we derive the underlying network of influence.

In this study, we required that compatible BNs verify *at least* the following dynamical properties:

- each observed transition (Figure 1b) is a transition possible in the asynchronous dynamics of the BN;
- the observed steady states correspond to fixed points of the BN: there must exist no outgoing transition from these configurations in the BN dynamics. The observed steady states are the five nodes in Figure 1b with ingoing edges but without outgoing edges, that is B, BP, P, T and no species ( $\_$ ).

In addition to the observations of the system, we investigated the impact of integrating expert knowledge on the structure of the influences. Specifically, in our case study, we considered *prior* trophic networks, which delineate possible interactions between species: we imposed that any BN must have its underlying interaction network be a *subgraph* of the prior network. The main effect of prior knowledge is to forbid interactions between species: if there is no edge from species A to B in the prior network, then the Boolean function of B must not depend directly on A.

Here, we considered three inference scenarios:

1. without prior knowledge on trophic interactions, except that the trophic network must be acyclic, which we refer to as the *Neutral* model;
2. only using interactions referenced in the (early) prior trophic network of Law et al. (2000); (Figure 1a), which we refer to as the *WLW Prior*;
3. only using interactions referenced in the extended prior trophic network from de Goër de Herve et al. (2022), which we refer to as the *GeA Prior*.

For the plankton ecosystem case study, we modelled the three (two observed and one theoretical) time series of Figure 2 as sequences of regimes following qualitative changes of bacteria and nitrate states: each observation is considered stable as long as the qualitative activity of bacteria and nitrate remains fixed. Then, once the condition changes, there must exist a trajectory toward the next observed regime. Moreover, the chosen prior network reflected expert knowledge on the possible as well as impossible trophic interactions between the different types and sizes of phytoplankton and zooplankton.

Integration of prior knowledge on trophic networks for inference of BNs requires a formal characterization of their translation in terms of BN influence graphs. To our knowledge, this link has never been addressed so far.

### 2.5.1 | From ecological networks to BN influence graphs

Trophic networks are directed graphs whose nodes are species and an edge from  $i$  to  $j$  means that  $i$  can be a predator of  $j$  and may feed on it. Moreover, the predation interaction may be qualified as *non-sufficient*, meaning that the prey alone is not sufficient to ensure the survival of the predator. For example, in the trophic network of Figure 1a, the species A cannot survive alone on species E, whereas it may survive alone on species T.

For each species, the BN defines a function that determines its next state, either present (true) or absent (false) depending on the current presence of all the other species. As defined in Section 2.3, the influence graph of a BN summarizes the effects of varying the presence state of individual species on the evaluation of the Boolean functions.

Given an ecological network, one can then translate the ecological interactions into potential influences between Boolean variables and Boolean functions. Let us denote by  $\mathcal{G}$  the potential influence graph of a BN corresponding to the ecological network. Essentially, if  $i$  is a predator for  $j$ , then we can deduce that:

- $f_i$  can depend positively on  $j$ , that is, there is positive influence from  $j$  to  $i$  ( $j \xrightarrow{+} i \in \mathcal{G}$ ) in its influence graph. This reflects that  $i$  may feed on  $j$ .
- $f_j$  can depend negatively on  $i$ , that is, there may be a negative influence from  $i$  to  $j$  ( $i \xrightarrow{-} j \in \mathcal{G}$ ) in its influence graph. This reflects that

$j$  can become absent because of  $i$ . More precisely, there may be a configuration where both  $i$  and  $j$  are present, which would then lead to  $j$  going extinct, while if  $i$  had not been there, it may survive.

In general, we allow positive self-loops on each species ( $i \xrightarrow{+} i \in \mathcal{G}$ ), independently to the presence of self-loops in the ecological network. This is notably necessary if the species  $i$  cannot shift from absent to present in the community, even in the presence of a prey  $j$ . In terms of Boolean function, this situation corresponds to the case whenever  $f_i(i = 1, j = 1) = 1$  ( $i$  sustains thanks to  $j$ ) and  $f_i(i = 0, j = 1) = 0$  ( $i$  cannot appear again even if  $j$  is present). There,  $f_i$  depends positively on  $i$  as the sole increase of  $i$  is sufficient to increase  $f_i$ .

The above rules enable to translate a trophic network into a BN influence graph  $\mathcal{G}$ , which can then be used to specify the inference problem: find a BN  $f$  using at most these influences ( $G(f) \subseteq \mathcal{G}$ ) and that is able to reproduce the observed transitions as close as possible. Figure 4 gives the translation of the WLW Prior as a BN influence graph.

It must be stressed that the obtained prior influence graph delimits the set of candidate BNs: a candidate BN must employ *at most* the influences of the prior influence graph. In particular, it may employ fewer influences while being able to reproduce the observed dynamics.

The non-sufficiency constraints can also be integrated as constraints on the Boolean function themselves. Indeed, if the prior knowledge indicates that species  $i$  cannot survive alone on species  $j$ , we impose that the Boolean function of  $i$  cannot be true on configurations where all species except  $i$  and  $j$  are in state 0.

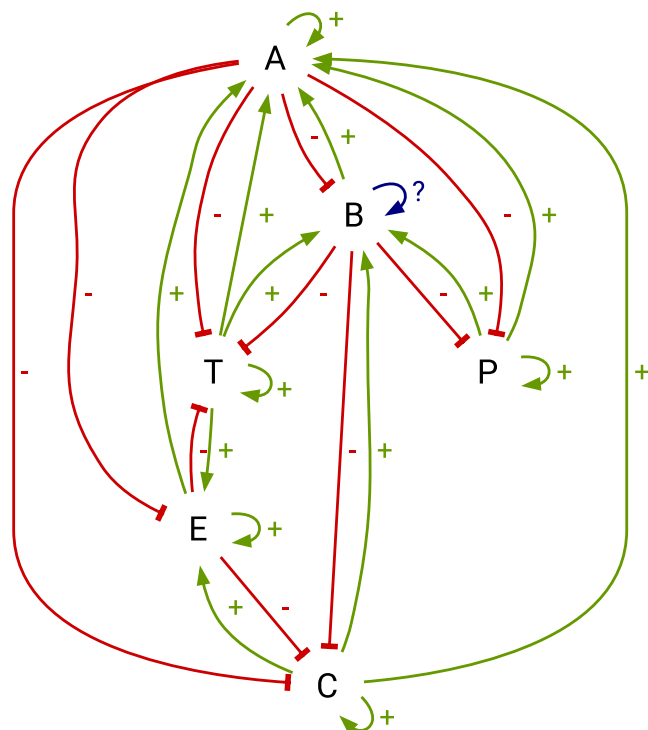


FIGURE 4 Influence graph corresponding to the WLW Prior (Figure 1a). Blue edges have an undetermined sign.

## 2.5.2 | From BNs to ecological networks

The Boolean functions of the BN may involve both trophic and competitions between species. The influence graph of a BN captures the signed causal relationships between the species and their Boolean activation functions. From these signed digraphs, we draw the corresponding ecological network as follows: There is an interaction from  $i$  to  $j$  in the ecological network if either  $f_i$  depends positively on  $j$  ( $j \xrightarrow{+} i \in G(f)$ ; 'presence of  $j$  can promote the presence of  $i$ ') or  $f_j$  depends negatively on  $i$  ( $i \xrightarrow{-} j \in G(f)$ ; 'presence of  $i$  can reduce the presence of  $j$ '). Thus, these rules capture both trophic and non-trophic relations, and do not permit distinguishing between them. Such a distinction would necessitate further inspection of the Boolean function themselves.

For instance, from the inspection of Boolean functions, one can distinguish sufficient from insufficient relations: An ecological interaction from species  $i$  to  $j$  is marked as sufficient (plain line in Figures 1 and 4) whenever it is sufficient that  $i$  and  $j$  are 1 in a state to satisfy the Boolean function associated with  $i$ , that is  $f_i(i = 1, j = 1) = 1$ . If the condition is not sufficient, we mark the interaction as insufficient (dashed line).

Finally, we may notice that the influence graph  $G(f)$  of the BN always includes more edges than the ecological network deduced with the above rules. Many cyclic influence graphs correspond to acyclic ecological networks. This is particularly the case for trophic interactions: predation relationships are typically reflected as both a positive edge from the prey to the predator and a negative edge from the predator to the prey.

Table 1 provides a summary of the main variables used in this paper.

## 3 | RESULTS

We present the results of the inference of BNs from the observed transitions of the protist system and the different levels of prior knowledge on the trophic network between the six species, and for the planktonic ecosystem from time-series data. In each case study, inferences have been run on a regular desktop computer (Intel Xeon 3.3Ghz with 64GB of RAM), and took between seconds to a couple of minutes depending on the cases.

### 3.1 | Protist case study

#### 3.1.1 | Inference of Boolean models of protist systems from zero knowledge

We inferred BNs that can reproduce at least the observed transitions of the protist system, without a priori on the ecological network, except that the latter must be acyclic. Therefore, any possible interaction is considered as soon as no species is directly or indirectly a prey of itself.

TABLE 1 Summary of main variables and technical terms used in this paper.

Term	Description
Boolean network theory	
BN	Boolean network
$f$	BN instance (Boolean map between binary vector of dimension $n$ )
$G(f)$	Influence graph of BN $f$
$i, j$	Variables denoting species indexes in a BN
Protist case study	
A	<i>Amoeba proteus</i>
B	<i>Blepharisma japonicum</i>
P	<i>Paramecium caudatum</i>
E	<i>Euplotes patella</i>
C	<i>Colpidium striatum</i>
T	<i>Tetrahymena pyriformis</i>
WLW Prior	a priori network from Law et al. (2000)
GeA Prior	a priori network from de Goër de Herve et al. (2022)
Neutral model	Sparsest BN reproducing input transitions
ZKM1, ZKM2, ZKM3	Zero Knowledge Models minimizing deviation from input transition
GeA best fit	BN from GeA prior having minimal deviation from input transitions
Plankton case study	
Bact	Bacteria
Nit	Nitrate
Proto	Protozoan
PicoP	Picoplankton
NanoP	Nanoplankton
MicroP	Microplankton
MicroZ	Microzooplankton
MesoZ	Mesozooplankton

We inferred the sparsest (neutral) model, which in our case was unique. It shows the minimal interactions that are needed to be able to reproduce (at least) the observations (Figure 5). It depicts trophic interactions already close to the WLW prior trophic network, with the notable addition of the predictions of T by B, C and P. The transition graph of the Neutral model (Figure 6) contains all the observed transitions with the same observed steady states, together with 81 additional transitions (deviation).

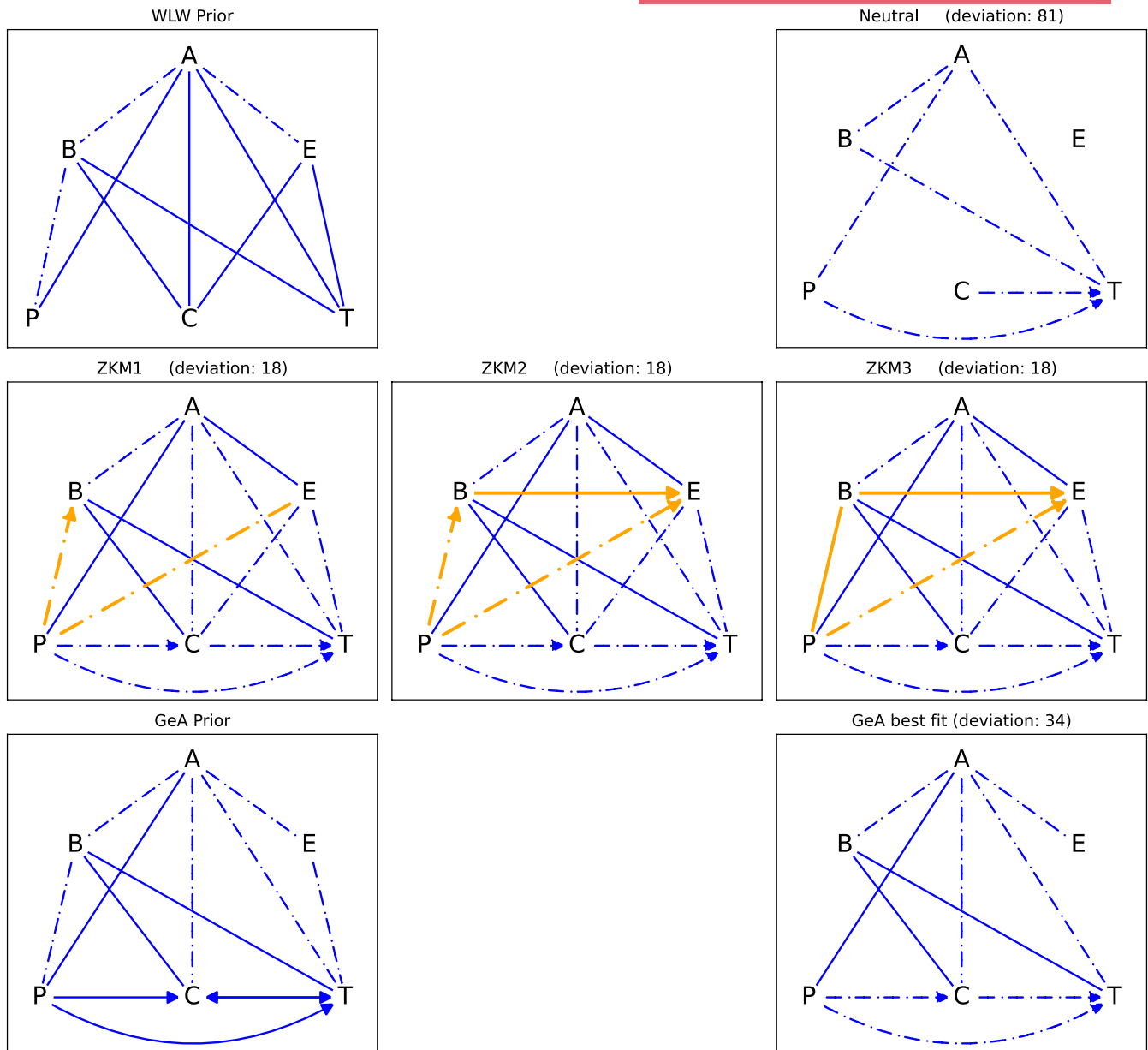
ZKM1/ZKM2/ZKM3 are alternative inferred models that minimize deviation while reproducing the observations and having an acyclic ecological network (Figure 5). Their dynamics include the observed transitions and 18 unobserved transitions (Figure 6). It is the minimum achievable with acyclic ecological interactions, observed transitions and steady states. The ZKM models differ only by the interactions between B, P and E. In particular, ZKM1 and ZKM2 involve predation of B by P, while ZKM3 needs the reverse. Compared to WLW Prior and Neutral, ZKM3 (and ZKM2) involve the predation of E by B and P. Considering each species size and their ecology, ZKM3 is the most realistic, as all interactions are compatible with a decreasing size from predators to prey.

Hence, even without a priori knowledge on the trophic interactions, the transitions and steady states are sufficient to identify plausible and sparse trophic networks by means of BN inference.

### 3.1.2 | Inference with expert knowledge from WLW Prior

The second inference was performed by constraining the ecological network to be contained within the WLW Prior (Figure 5), including indicated non-sufficiency of interactions. We ignored constraints related to the sufficiency of trophic interactions to maintain a species in order to consider more candidate models.

It resulted that *no model* using exclusively those pre-determined ecological interactions can reproduce all the observed transitions and steady states. The reason lies in the fact that, in the trophic network, T is not related to C and P. However, from the observations, T can survive alone in steady state, but in the presence of C or P, T goes extinct. Indeed, the following transitions are given (Figure 1b):



**FIGURE 5** Prior and inferred ecological interaction networks. Unless specified, the edge direction is towards the lower; there are no edges with default direction overlapped by a directed edge. Dashed-dotted edges represent predation that is not sufficient for the predator to survive only with this prey. Bacteria are not displayed.

- $CT \rightarrow C$
- $CPT \rightarrow CP$
- $PT \rightarrow P$

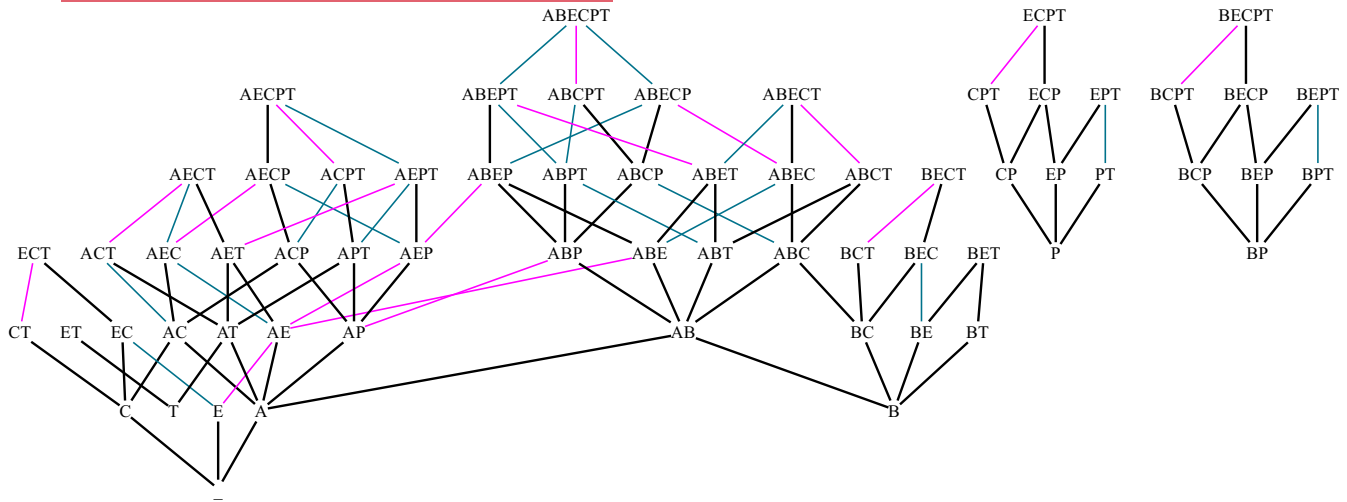
It is therefore impossible to explain these above transitions without introducing additional interactions, thus horizontal predations or competitions. Note that this was captured by the Neutral and ZKM models.

### 3.1.3 | Extended WLW Prior

de Goër de Herve et al. (2022) proposed an extended trophic network, coupled with potential competition, reproduced as 'GeA Prior'

in Figure 5. The GeA Prior network contains a cycle between species C and T to indicate potential competition or preference relations. As we explicitly imposed that admissible models have an acyclic ecological network, at most one direction can be employed in solution Boolean networks.

From this extended ecological network, we automatically identified one (unique) model, minimizing its deviation from observations. This model, whose trophic network is shown in Figure 5 as 'GoA best fit', introduces 34 unobserved transitions (Figure 6). It is the best fit model while using only the interactions listed in the GeA Prior. Its deviation is greater than that of the ZKM models. Compared to ZKM3, the model misses numerous interactions involving E, and the predation of P by B. Moreover, the predation of E is sufficient for A



**FIGURE 6** Transition graphs of inferred models. Transitions in black plain lines are observations and are shared by all the models; in magenta are the 18 additional transitions of the ZKM3 and GoA models; in (dashed) teal are 16 additional transitions of the GoA model. Transitions of the Neutral, which include GoA transitions, are provided in [Figure S1](#).

to sustain in ZKM models, while it is declared as insufficient in the GoA best fit model.

This model shows a trophic network rather close to the WLW prior, except for the species T, which is here predated by P and C, which enables reproducing the observed transitions related to the extinction of T in their presence. Hence, again, the absence of such trophic interactions made the inference of a compatible model using only WLW prior impossible.

### 3.1.4 | Plankton ecosystem case study

In this case study with eight variables, the observations ([Figure 2](#)) specify only a fraction of the dynamics (in total, the existence of trajectories between seven time steps). Hence, one can expect numerous diverse candidate models, despite a rather constrained prior knowledge. From the prior influence graph ([Figure 7a](#)), we identified 2400 different minimal influence graphs: no pair of these influence graphs has inclusion relationships, but any solution Boolean network has influence graphs which include at least one of those 2400.

[Figure 7](#) shows the intersection and union between these influence graphs and the corresponding ecological networks.

## 4 | DISCUSSION

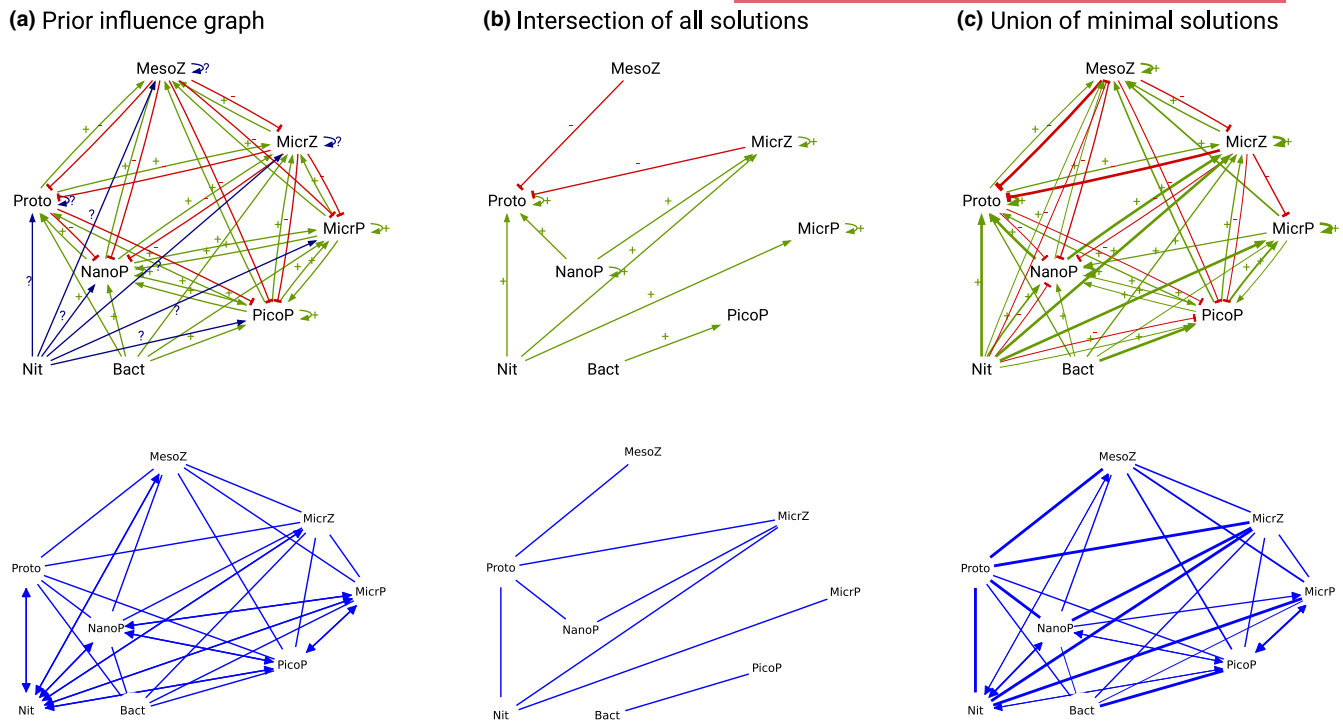
### 4.1 | Protist case study

The WLW Prior, and Neutral and GeA model present interactions more or less vertical, in terms of ecology and body sizes, leaving more or less room to the non-trophic (e.g. competition and other horizontal) interactions. Furthermore, some models obtained without a priori on the ecological network appear unrealistic (ZKM1 and ZKM2) as they infer 'reverse predations' (smaller preys possibly

eating larger predators). In aquatic ecology, the size of the mouth of a predator is a typical constraint for the size of the prey. The reverse situation might happen in the case that prey adults sometimes feed on predator juveniles (Choh et al., 2012). Here, we often observed that the reverse role of the E species, which is not so surprising considering only its body size comparable to the P species (Warren et al., 2006; Weatherby et al., 1998). For this reason, this might be experimentally verified. We also observed here that T and E species often appear to be discriminant (ZKMs) and should be studied.

Our explorations also confirm that one single interaction may drastically change the computed dynamics, but not necessarily the deviation to observations, as in the ZKM models. Similarly, we note here the importance of what could be called the 'interaction status', namely the ability to sustain the predator species (plain edges) or not (dotted edges), and the fact that some prey may be preferred or not (among several) by a predator species is crucial (e.g., ZKM2, species A). It might be carefully verified experimentally, as it may change the fate of the community by systematically removing a species before another one (and so forth) (de Goër de Herve et al., 2022; Fukami, 2015). For this reason, too, the model deviation itself should not be regarded as the single discriminant criterion of inference, as expert knowledge (on species behaviour) may sharply change the inferred network, too. And one (carefully selected) transition would be sufficient to refute an inferred model, which is reassuring.

All the models we infer are able, by design, to mimic all the observed transitions, including the steady states. The minimum number of additional transitions, we qualified as deviation, depends on the a priori on the network structure. One can observe in [Section 3](#) (and in [Figure S1](#)) that most of the additional transitions form so-called diamonds of transitions. For example, starting from BEPT, the observations only relate transitions from BEPT to BEP to BP, and BPT to BP. The ZKM3 model adds the transition from BEPT to BPT. In qualitative dynamics, these diamonds typically reflect *concurrent*



**FIGURE 7** (a) Prior influence graph used to constrain the Boolean network inference, (b) intersection of the influence graphs computed from all solution models, (c) union of the influence graphs of the 2400 minimal models. Edge thickness indicates frequency. The edges of the intersection (b) have the highest frequency and are present in all the models. The second row displays the ecological networks derived from the corresponding influence graphs.

events: From BEPT, species E and T eventually disappear. If these events are independent of each other, they can disappear in any order, generating diamond-shaped transitions. As the asynchronous dynamics of BNs do not capture features related to speed or priority, one can indeed expect that there is not enough information to enforce a specific ordering in species disappearance. The only possibility of having more controlled transitions, that is having a more complex causal structure to explain that a transition is possible but not another from the same state. Thus, the importance of the prior ecological network, which, by denying direct interactions (those not referenced in the prior network), prevents deriving aberrant causal relations to fit as much as possible with the data.

## 4.2 | Planktonic ecosystem case study

Overall, the union of inferred networks reveals interesting and expected interactions between species (Figure 7c). The most frequent interactions (in bold) are often correct, yet with a few anomalies discussed below. Nitrates are dissolved molecules that are consumed by autotroph species only (i.e. all types of phytoplanktons). Hence, MicrZ and Proto, as heterotroph species, cannot ingest such inorganic molecules, and these bold interactions are wrongly inferred. The self-loops on some variables, meaning that those species can survive alone (on the bacterial resource only), should concern only basal and autotroph species and, thus, are inappropriate for heterotroph species such as Proto, MicrZ and MesoZ. Finally, MicrZ can

indeed eat NanoP, as many other interactions suggested by this network.

Remarks are relatively similar for the intersection (and minimal) network (Figure 7b), with a few heterotroph species sometimes eating inorganic molecules (MicrZ and Proto), and zooplankton, which would starve on bacterial resource only (MicrZ and MesoZ). As a conclusion for this more realistic inference test, our method was able to infer the most relevant species interactions, considering the few constraints that we have chosen (Figure 7a). It would have been possible to infer a more precise network with a larger set of constraints on interactions (e.g. between autotroph and/or heterotroph species), but this would have required a more extended knowledge of the system. Conversely, it would have been much more efficient to increase the number of observed states (at stations), that is to refine the pathways used for training the inference model, for better constraining the trophic network inferred.

## 4.3 | Inference methodology

We demonstrated here an interactive method for generating hypotheses on underlying mechanisms driving ecological systems by means of inference of qualitative models. Automatically inferring (without or with little expert knowledge) ecological interaction networks becomes today mandatory in many observed ecosystems (Milns et al., 2010; Pocock et al., 2021). Our qualitative and possibilistic method (i.e. asynchronously exploring all possible transitions)

appears here capable of identifying several interaction networks responsible for observed dynamics, among which some are minimalist and display low deviations between dynamical observations and computations. In this study, we showed a complementary and parsimonious approach, which requires no quantitative data and little expert knowledge about the system studied (Gaucherel, Fayolle, et al., 2024; Gaucherel & Pommereau, 2019; Paulevé et al., 2020). This enables one to efficiently test several interaction networks fulfilling the observed dynamics and derive candidate models that can then serve as a basis to derive more fine-grained and quantitative models.

Without specific prior knowledge of the trophic network, our approach automatically suggests the simplest explanation for the observations, that is using the least trophic interactions or least transitions as possible. This can constitute a first working hypothesis on the underlying ecological model, without the need for quantification. The inferred trophic interaction may then be validated or refuted, possibly by additional experiments and observations. This new expert knowledge can then be integrated, and a new inference process will return new models satisfying these additional constraints, in an iterative procedure (Auclair et al., 2017; Baldan et al., 2018). Our inference framework can also conclude the absence of models satisfying both desired dynamical and structural properties. This typically indicates that prior knowledge or qualitative interpretation of the data must be revised to obtain a comprehensive logical explanation of the system dynamics.

## 4.4 | Theoretical guarantees, applicability, and limitations of the approach

### 4.4.1 | Advantages and limitations of Boolean modelling

Boolean networks have a long and successful history of applications in systems biology, notably for modelling gene expression dynamics at the cellular level (Collombet et al., 2017; Montagud et al., 2022; Thomas, 1973; Zañudo et al., 2021). As for ecological systems, the abundance of species (the expression of genes in cellular modelling) is, by nature, quantitative. Nevertheless, qualitative modelling is a common practice to tackle such complex systems, and provides several advantages.

The prime and foremost reason for Boolean modelling adoption in systems biology is their *easier parametrization* compared with quantitative models such as ordinary differential equations. Boolean networks require few parameters, and those parameters are of discrete and finite nature. This makes the modelling (automatic, as proposed in this paper, or manual) much more tractable and robust to noise in observations.

Another direct benefit is the ability to encompass many components, often of highly different natures: In the literature in systems biology, Boolean models are composed of several dozen up to a

few hundreds of genes and/or molecules (Abou-Jaoudé et al., 2015; Montagud et al., 2022).

By nature, Boolean networks are qualitative, making them little practical to reason on quantitative features of the modelled systems. Because a species is either absent or present, this enforces a drastic coarse-graining of the system dynamics, which, depending on the application, may become irrelevant. In many cases, Boolean networks, and thus interaction networks here, can be seen as a 'backbone' model, which integrates interactions between species that drive the overall dynamics (Gaucherel, Cosme, et al., 2024). A concrete example in systems biology is the modelling of gene and signalling networks, where Boolean networks are able to predict the effect of gene mutations (forcing a gene to be silenced or, on the contrary, to be activated) (Flobak et al., 2015; Zañudo et al., 2021). Boolean models can then be refined into quantitative models by integrating kinetics and finer description of species states (Fages & Soliman, 2008; Paulevé et al., 2020; Pratapa et al., 2020). In particular, any system component usually can be 'unfolded' (or folded) into a finer interaction network of its own components (Gaucherel, Fayolle, et al., 2024).

### 4.4.2 | Exhaustiveness of logic-based search for Boolean networks

The model inference methodology we present in this paper relies on logic-based search of Boolean networks satisfying given dynamical and structural properties. We used the BoNESIS software (Chevalier, Boyenval, et al., 2024), which employs logic programming and Boolean satisfiability technologies to accomplish this search. The main principle is that the model inference is encoded as a Boolean equation, which integrates both the desired dynamical and structural properties, and where variables correspond to the Boolean function to identify. Then, a solution to the Boolean equation corresponds to a Boolean network that is guaranteed to satisfy the specification (Chevalier et al., 2019). As mentioned in Section 2.4, BoNESIS relies on Answer-Set Programming and CLINGO solver. The solving algorithms allow performing an exhaustive reasoning on solutions. Therefore, if no solutions are returned, it means that the problem has no solution, that is there is no model satisfying the input specification. An opposite (and usual) case is whenever there is a very large number of solutions, preventing their exhaustive enumeration. In such cases, logic programming and CLINGO allow identifying the *minimal* models, which are usually much fewer, as well as obtaining the intersection and union over all models. Note that computing union and intersection networks can be done without an exhaustive enumeration, and is usually very efficient (Figure 7).

In the presented case studies, we exploited this exhaustive reasoning capability to extract minimal models, that is models that are included in any other solution models, and thus include necessary interactions to be able to reproduce observed behaviour.

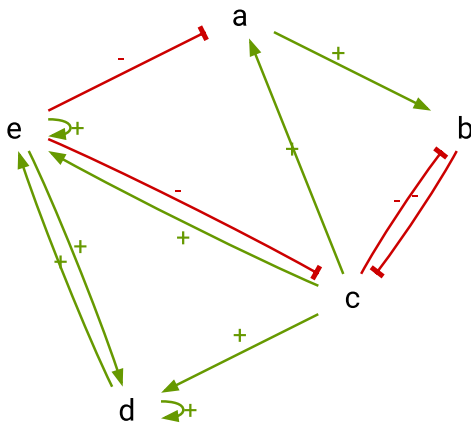
#### 4.4.3 | Impact of the nature and number of constraints on minimal models

With BoNESIS, the set of candidate Boolean networks is primarily delimited by a given prior influence graph: Any Boolean network whose influence graph is contained in it will be considered as a potential solution. Then, provided additional constraints, either related to the dynamics (e.g. existence of specific steady states or transitions), or to the model structure (e.g., acyclicity of the tropic network, presence or absence of a specific direct regulation), shrink the admissible solutions. As mentioned in the previous paragraph, ASP enables one to efficiently browse the minimal models, here minimal influence graphs: an influence graph of a Boolean network is minimal if no other Boolean network has an influence graph being a strict subgraph of the former. Thus, one minimal model can be seen as representative of a large set of solutions, those including at least the influences of the minimal model. As showcased in the applications, there can be several distinct minimal models. Moreover, while the total number of solutions necessarily decreases when the number of dynamical or structural constraints increases, the number of minimal models is usually not monotone with the addition of constraints.

We illustrate this observation and how the number of constraints can impact the number of minimal models using synthetic models generated randomly. We built four models on the basis of five variables, resulting in the one in Figure 8 and the three others depicted in the online Supporting Information (Paulevé & Gaucherel, 2025), with similar behaviours. Such synthetic examples only show specific instances of the effect of constraints on the number of (minimal) models and their capability to recapitulate interactions and dynamics of the ground truth model, which are all heavily case-dependent.

The asynchronous dynamics of the synthetic model contains two steady states and 64 transitions. Its influence graph contains 12 direct interactions. Starting from the complete (cyclic) graph, we studied how the number of minimal models varies when increasing the number of steady states or transitions constraints. For the impact of steady states (Figure 8b), we performed the inference of minimal models with specifying zero, one (randomly selected), and the two steady states, and the 64 asynchronous transitions. For the impact of transitions (Figure 8c), we performed the inference by progressively specifying more and more transitions (randomly ordered), while always imposing the two steady states. One can observe that the number of minimal models can

(a) influence graph of synthetic model



(b) impact of varying number of input steady states

min models	size	overlap (/12)	spurious trans	missing trans
0	16	4-7	0-7	16-80
1	2	8	7-8	16-24
2	1	8	8	16

(c) impact of varying number of input transitions

min models	size	overlap (/12)	spurious trans	missing trans
0	256	4	0-4	4-48
4	48	4	1-4	12-44
8	240	6-8	2-6	8-43
12	60	6-7	2-6	12-43
16	15	6-7	3-6	24-39
20	5	6-7	4-6	24-31
24	5	6-7	4-6	24-31
28	6	7-9	5-7	20-36
32	18	8-12	5-8	16-39
36	12	8-13	5-8	16-41
40	8	8-14	5-8	16-43
44	8	8-14	5-8	16-43
48	2	8-10	7-8	16-23
52	2	8-10	7-8	16-23
56	2	8-10	7-8	16-23
60	1	8	8	16
64	1	8	8	16

FIGURE 8 Synthetic example for demonstrating the effect of the number of constraints on the number of inferred minimal models.

The synthetic Boolean network was randomly generated and has functions  $f_a(\mathbf{x}) = \mathbf{x}_c \wedge \neg \mathbf{x}_e$ ,  $f_b(\mathbf{x}) = \mathbf{x}_a \wedge \neg \mathbf{x}_c$ ,  $f_c(\mathbf{x}) = \neg \mathbf{x}_b \wedge \neg \mathbf{x}_e$ ,  $f_d(\mathbf{x}) = \mathbf{x}_e \vee (\mathbf{x}_c \wedge \mathbf{x}_d)$ ,  $f_e(\mathbf{x}) = \mathbf{x}_e \vee (\mathbf{x}_c \wedge \mathbf{x}_d)$ . Its influence graph is represented in (a). Its asynchronous dynamics possesses two steady states and 64 transitions. (b) Effect of varying the number of steady states in the inference specification, while ensuring the 64 transitions. The table reports the number of minimal models (influence graphs), their size range and overlap with ground truth (a), and number of spurious and missing asynchronous transitions. (c) Effect of varying the number of transitions in the inference specification, while ensuring the two steady states.

both increase and decrease while adding more constraints, which reflects the change of shape of the global solution space. As expected, the more constraints we specify, the closer the minimal models are to the ground truth, both in terms of recall of the influence graph (overlap) and in terms of recall and precision of asynchronous transitions. In that particular case, when specifying the 64 transitions, the minimal models are smaller than the ground truth and generate spurious transitions. This is because we did not specify that the set of transitions is complete and, thus, the inference considers any model that can reproduce *at least* the 64 transitions. Another remark is that the number of minimal solutions is usually very small compared to the total number of solutions. For instance, for the inference with the 64 transitions but with no steady state constraints, there are 16 minimal models, whereas a partial enumeration provided more than 628 million models.

#### 4.4.4 | Scalability of logic-based model search

The case studies we tackled in this paper can be considered of modest scale. All the computations have been performed in a few seconds or minutes. The tool BoNESis has been successfully applied to infer models of cellular regulation networks comprising thousands of genes (Chevalier, Becker, et al., 2024), showing a potential large scalability. Technically, the scalability of our logic-based approach is tied to two factors: the size of the logic program to be solved, and the solving itself (Chevalier et al., 2019). The source of complexity in our approach is the potential exponential size of the Boolean functions to identify: The Boolean function of a species  $i$  can be of size exponential with the number of species that influence it. In other words, the in-degree of the influence graph is a crucial feature. Yet, Chevalier, Becker, et al. (2024) have been able to address cases where genes can be influenced by hundreds of other genes. In these cases, however, we have to manually cap the complexity of the Boolean function to infer. The downside is that some admissible models, although having complex functions, will be excluded from the search. Having generated the logic program, the search of solutions is then similar to the infamous SAT (Boolean satisfiability) problem, which is NP-complete in general. Nevertheless, modern solvers such as CLINGO are able to address large problems, with several hundreds of thousands of variables. This is however quite unpredictable: some problems with a few hundreds of variables can be much harder to solve than some extremely large problem (Ganesh & Vardi, 2020).

#### 4.4.5 | Pitfalls and limitations of Boolean network inference

Whereas BoNESis provides a language for specifying complex dynamical properties the Boolean networks must have, one of the main pitfalls is the specification of the state of the species. In the case studies of this paper, the observations were already 'binarized'

(i.e. qualified (Gaucherel & Pommereau, 2019)). However, observations are often quantitative. It is then necessary to transform these data into qualitative Boolean states, possibly with undetermined values. Binarization can be achieved by devising thresholds, or employing a statistical model to classify quantitative states into either absent, present or undetermined states (Magaña López et al., 2024). This step can be very sensitive, as it impacts the properties from which the model is built.

Another regular pitfall of such inference method is its scalability, as discussed in the previous paragraph. As the model inference method aims at identifying models that formally fulfil all the input constraints, such a model may not exist, or be very difficult to identify. In the case of absence of solution, one classical approach is to relax some hard constraints into optimization objectives: find models that satisfy as much dynamical properties as possible. When facing instances where the search takes a very long time, one can also relax some dynamical constraints and reduce the number of components to enhance tractability. The opposite case is whenever the number of solutions is very large, essentially due to an under-specified inference problem. This situation is, however, much more practical to handle: One can focus on minimal models, compute intersections, or project solutions on a subset of components.

The case whenever there are too many solutions is one of the main limitations of logic-based model inference, while also being an indicator of a lack of observations to constrain the inference problem. Indeed, by design, the output models are guaranteed to possess the specified dynamical properties. Thus, in the case that there are a few solutions, it is guaranteed that the 'true' (i.e., searched for) model is one of them. However, when facing too many solutions, one has to enumerate only a subset of them (for example, the minimal ones). Then, the 'true' model is not guaranteed to be in this subset of solutions. A major breakthrough would be the capability to optimize the predictive power of output solutions. In the current settings, output models must be challenged by expert knowledge and further observations in order to finely evaluate their relevance.

#### 4.4.6 | Application to other case studies

In order to apply the BoNESis inference method efficiently to a dataset, one needs to delineate the *domain of candidate models* and to formalize a *qualitative and dynamical interpretation of the data*. The domain of candidate models is typically specified using a prior knowledge influence graph from which possible Boolean networks will be derived. As illustrated in this manuscript, the prior knowledge can reflect expert knowledge on possible and impossible interactions between species. Data typically include (partial) observations of species at different times and conditions. In the case, these data are quantitative, a binarization must be performed, as addressed in the previous subsection. Then, the observations must be translated into dynamical properties that a valid model has to possess: for example, existence of a trajectory leading from one observation to another, stability of a given

observation. Such constraints are expressed in BoNESis by means of a specific language (see <https://bnediction.github.io/bonesis/> for documentations and examples). This modelling step is the heart of the inference methodology, as it formally specifies which models are considered admissible.

Then, BoNESis provides different methods for accessing the solution space, for instance by enumerating solution Boolean networks, or enumerating their influence graphs. By analysing the solutions, it is usual to identify models that present features that are not desired, according to expert knowledge, for instance. This can lead to refining the domain of candidate models, for example by denying some aberrant interactions, or including additional constraints on the dynamics. The above steps must then be repeated with the revised inference specification.

Future developments could provide additional features for expressing further constraints on the dynamics of Boolean networks, as well as for analysing ensembles of inferred solutions. Besides clustering admissible models, as performed by (Chevalier, Becker, et al., 2024), one may deduce observations or experiments that would allow discriminating a large set of models, and thus reduce model uncertainty.

## 5 | CONCLUSIONS

The identification of species interaction networks at play in the field, together with the inference of related dynamical models, is a major challenge for the ecological community and further (Fukami, 2015; Serván & Allesina, 2021). This discipline is on the way to developing powerful inference, often based on frequentist and Bayesian models (Milns et al., 2010; Pocock et al., 2021). Yet, they often require a large amount of data and expert fine-tuning rarely available in ecology and environmental sciences.

The qualitative modelling and inference approaches, such as the one presented in this paper, bring the capability to generate models compatible with both observations and prior expert knowledge. Due to the discrete and coarse-grained abstraction, inference methods are able to extract minimal models, and more specifically interactions that must be present to reproduce the observed behaviours (Paulevé et al., 2020). However, due to the coarse-graining imposed by Boolean modelling, such models can be limited in the prediction of certain quantitative features of ecological systems.

Therefore, a promising perspective would be to combine purely qualitative approaches to delineate a first ensemble of candidate models based on logical reasoning, before employing fine-grained modelling approaches to incorporate more information related to temporized time and species abundances.

### AUTHOR CONTRIBUTIONS

Loïc Paulevé and Cédric Gaucherel conceived the ideas and designed methodology; Loïc Paulevé performed computational experimentation and Cédric Gaucherel performed analysis; Loïc Paulevé and Cédric Gaucherel wrote the manuscript.

### ACKNOWLEDGEMENTS

We warmly thank Mathieu de Goër de Herve who developed the first model from the discrete-event family. We also thank Franck Pommereau and Christine Dupuy for fruitful discussions on early versions of this work. LP was supported by the French Agence Nationale pour la Recherche (ANR) in the scope of the projects 'BNeDiction' (grant number ANR-20-CE45-0001) and 'REBON' (grant number ANR-23-CE45-0008).

### CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

### PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.70090>.

### DATA AVAILABILITY STATEMENT

Data and code for reproducing the protists and plankton case studies, as well as the synthetic model experiments, are available at <https://github.com/pauleve/bonesis-ecology-casestudies> and archived at <https://doi.org/10.5281/zenodo.15292813> (Paulevé & Gaucherel, 2025). The inference has been performed using BoNESis v0.6.5, available at <https://bnediction.github.io/bonesis> and archived at <https://doi.org/10.5281/zenodo.10891199> (Paulevé, 2024).

### ORCID

Loïc Paulevé  <https://orcid.org/0000-0002-7219-2027>

Cédric Gaucherel  <https://orcid.org/0000-0002-4521-8914>

### REFERENCES

- Abou-Jaoudé, W., Monteiro, P. T., Naldi, A., Grandclaoudon, M., Soumelis, V., Chaouiya, C., & Thieffry, D. (2015). Model checking to assess t-helper cell plasticity. *Frontiers in Bioengineering and Biotechnology*, 2, 86. <https://doi.org/10.3389/fbioe.2014.00086>
- Auclair, E., Peyrard, N., & Sabbadin, R. (2017). Apprentissage de réseau bayésien dynamique étiqueté avec connaissance a priori sur la structure du réseau. In *49èmes Journées de Statistique*. INRAE. <https://hal.inrae.fr/hal-02940468>
- Baldan, P., Bocci, M., Brigolin, D., Cocco, N., Heiner, M., & Simeoni, M. (2018). Petri nets for modelling and analysing trophic networks. *Fundamenta Informaticae*, 160(1–2), 27–52. <https://doi.org/10.3233/fi-2018-1673>
- Baral, C. (2003). *Knowledge representation, reasoning and declarative problem solving*. Cambridge University Press.
- Barroso-Bergadà, D., Pauvert, C., Vallance, J., Delière, L., Bohan, D. A., Buée, M., & Vacher, C. (2020). Microbial networks inferred from environmental DNA data for biomonitoring ecosystem change: Strengths and pitfalls. *Molecular Ecology Resources*, 21(3), 762–780. <https://doi.org/10.1111/1755-0998.13302>
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer-Verlag.
- Chevalier, S., Becker, J., Gui, Y., Noël, V., Su, C., Jung, S., Calzone, L., Zinovjev, A., del Sol, A., Pang, J., Sinkkonen, L., Sauter, T., & Paulevé, L. (2024). Data-driven inference of Boolean networks from transcriptomes to predict cellular differentiation and reprogramming. *bioRxiv*. <https://doi.org/10.1101/2024.10.21.618706>

- Chevalier, S., Boyenval, D., Magaña López, G., Roncalli, T., Vaginay, A., & Paulevé, L. (2024). BoNesis: A python-based declarative environment for the verification, reprogramming, and synthesis of Most Permissive Boolean networks. In *22th international conference on Computational Methods in Systems Biology (CMSB 2024)*, LNCS. Springer. <https://hal.science/hal-04629083/file/main.pdf>
- Chevalier, S., Froidevaux, C., Paulevé, L., & Zinovyev, A. (2019). Synthesis of Boolean networks from biological dynamical constraints using answer-set programming. In *2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI)* (pp. 34–41). IEEE. <https://doi.org/10.1109/ICTAI.2019.00014>
- Choh, Y., Ignacio, M., Sabelis, M. W., & Janssen, A. (2012). Predator-prey role reversals, juvenile experience and adult antipredator behaviour. *Scientific Reports*, 2, 728. <https://doi.org/10.1038/srep00728>
- Collombet, S., van Oevelen, C., Sardina Ortega, J. L., Abou-Jaoudé, W., Di Stefano, B., Thomas-Chollier, M., Graf, T., & Thieffry, D. (2017). Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 114(23), 5792–5799. <https://doi.org/10.1073/pnas.1610622114>
- Dambacher, J. M., Luh, H.-K., Li, H. W., & Rossignol, P. A. (2003). Qualitative stability and ambiguity in model ecosystems. *The American Naturalist*, 161, 876–888. <https://doi.org/10.1086/367590>
- de Goër de Herve, M., Thomas, C., Cosme, M., Warren, P., & Gaucherel, C. (2022). Is a community state reachable, and why? *TechRxiv*. <https://doi.org/10.22541/au.164301558.85072720/v1>
- Fages, F., & Soliman, S. (2008). Abstract interpretation and types for systems biology. *Theoretical Computer Science*, 403(1), 52–70. <https://doi.org/10.1016/j.tcs.2008.04.024>
- Flobak, Ø. A., Baudot, A., Remy, E., Thommesen, L., Thieffry, D., Kuiper, M., & Lægreid, A. (2015). Discovery of drug synergies in gastric cancer cells predicted by logical modeling. *PLoS Computational Biology*, 11(8), e1004426. <https://doi.org/10.1371/journal.pcbi.1004426>
- Fukami, T. (2015). Historical contingency in community assembly: Integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 1–23. <https://doi.org/10.1146/annurev-ecolsys-110411-160340>
- Ganesh, V., & Vardi, M. Y. (2020). *On the unreasonable effectiveness of SAT solvers*. Cambridge University Press.
- Gaucherel, C., Cosme, M., Nous, C., & Franck, P. (2024). A single changing hypernetwork to represent (social)-ecological dynamics. *Peer Community Journal*, 4, e104. <https://doi.org/10.24072/pcjournal.482>
- Gaucherel, C., Fayolle, S., Savelli, R., Philippine, O., Pommereau, F., & Dupuy, C. (2024). Diagnosis of planktonic trophic network dynamics with sharp qualitative changes. *Peer Community Journal*, 4, e58. <https://doi.org/10.24072/pcjournal.417>
- Gaucherel, C., & Pommereau, P. (2019). Using discrete systems to exhaustively characterize the dynamics of an integrated ecosystem. *Methods in Ecology and Evolution*, 10, 1615–1627. <https://doi.org/10.1111/2041-210X.13242>
- Gebser, M., Kaminski, R., Kaufmann, B., & Schaub, T. (2014). Clingo = ASP + control: Preliminary report. *CoRR*, abs/1405.3694. <https://doi.org/10.48550/arXiv.1405.3694>
- Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly connected nets. *Journal of Theoretical Biology*, 22, 437–467. [https://doi.org/10.1016/0022-5193\(69\)90015-0](https://doi.org/10.1016/0022-5193(69)90015-0)
- Law, R., Weatherby, A. J., & Warren, P. H. (2000). On the invasibility of persistent protist communities. *Oikos*, 88(2), 319–326. <https://doi.org/10.1034/j.1600-0706.2000.880210.x>
- Magaña López, G., Calzone, L., Zinovyev, A., & Paulevé, L. (2024). sc-BoolSeq: Linking scRNA-seq statistics and Boolean dynamics. *PLoS Computational Biology*, 20(7), 1–25. <https://doi.org/10.1371/journal.pcbi.1011620>
- Masclaux, H., Tortajada, S., Philippine, O., Robin, F.-X., & Dupuy, C. (2014). Planktonic food web structure and dynamics in freshwater marshes after a lock closing in early spring. *Aquatic Sciences*, 77(1), 115–128. <https://doi.org/10.1007/s00027-014-0376-1>
- Milns, I., Beale, C. M., & Smith, V. A. (2010). Revealing ecological networks using bayesian network inference algorithms. *Ecology*, 91(7), 1892–1899. <https://doi.org/10.1890/09-0731.1>
- Montagud, A., Béal, J., Tobalina, L., Traynard, P., Subramanian, V., Szalai, B., Alföldi, R., Puskás, L., Valencia, A., Barillot, E., Saez-Rodriguez, J., & Calzone, L. (2022). Patient-specific boolean models of signalling networks guide personalised treatments. *eLife*, 11, e72626. <https://doi.org/10.7554/eLife.72626>
- Paulevé, L. (2024). bnediction/bonesis: v0.6.5. *Zenodo*. <https://doi.org/10.5281/zenodo.10891199>
- Paulevé, L., & Gaucherel, C. (2025). Notebooks for article “inference of ecological networks and possibilistic dynamics based on boolean networks from observations and prior knowledge”. *Zenodo*. <https://doi.org/10.5281/zenodo.15292814>
- Paulevé, L., Kolčák, J., Chatain, T., & Haar, S. (2020). Reconciling qualitative, abstract, and scalable modeling of biological networks. *Nature Communications*, 11(1), 4256. <https://doi.org/10.1038/s41467-020-18112-5>
- Pocock, M. J. O., Schmucki, R., & Bohan, D. A. (2021). Inferring species interactions from ecological survey data: A mechanistic approach to predict quantitative food webs of seed feeding by carabid beetles. *Ecology and Evolution*, 11(18), 12858–12871. <https://doi.org/10.1002/ece3.8032>
- Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A., & Murali, T. M. (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature Methods*, 17(2), 147–154. <https://doi.org/10.1038/s41592-019-0690-6>
- Serván, C. A., & Allesina, S. (2021). Tractable models of ecological assembly. *Ecology Letters*, 24(5), 1029–1037. <https://doi.org/10.1111/ele.13702>
- Thomas, R. (1973). Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42(3), 563–585. [https://doi.org/10.1016/0022-5193\(73\)90247-6](https://doi.org/10.1016/0022-5193(73)90247-6)
- Veloz, T. (2019). The complexity–stability debate, chemical organization theory, and the identification of non-classical structures in ecology. *Foundations of Science*, 25(1), 259–273. <https://doi.org/10.1007/s10699-019-09639-y>
- Warren, P. H., Law, R., & Weatherby, A. J. (2003). Mapping the assembly of protist communities in microcosms. *Ecology*, 84(4), 1001–1011. [https://doi.org/10.1890/0012-9658\(2003\)084\[1001:mtaopc\]2.0.co;2](https://doi.org/10.1890/0012-9658(2003)084[1001:mtaopc]2.0.co;2)
- Warren, P. H., Law, R., & Weatherby, A. J. (2006). Invasion biology as a community process: Messages from microbial microcosms. In *Conceptual ecology and invasion biology: Reciprocal approaches to nature* (pp. 343–367). Kluwer Academic Publishers. [https://doi.org/10.1007/1-4020-4925-0\\_16](https://doi.org/10.1007/1-4020-4925-0_16)
- Weatherby, A. J., Warren, P. H., & Law, R. (1998). Coexistence and collapse: An experimental investigation of the persistent communities of a protist species pool. *Journal of Animal Ecology*, 67(4), 554–566. <https://doi.org/10.1046/j.1365-2656.1998.00212.x>
- Werner, E. E., & Peacor, S. D. (2003). A review of trait-mediated indirect interactions in ecological communities. *Ecology*, 84(5), 1083–1100. [https://doi.org/10.1890/0012-9658\(2003\)084\[1083:AROTII\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[1083:AROTII]2.0.CO;2)
- Yu, J., Smith, V. A., Wang, P. P., Hartemink, A. J., & Jarvis, E. D. (2004). Advances to bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*, 20(18), 3594–3603. <https://doi.org/10.1093/bioinformatics/bth448>
- Zañudo, J. G. T., Mao, P., Alcon, C., Kowalski, K., Johnson, G. N., Xu, G., Baselga, J., Scaltriti, M., Letai, A., Montero, J., Albert, R., & Wagle, N. (2021). Cell line-specific network models of ER+ breast cancer identify potential PI3k $\alpha$  inhibitor resistance mechanisms and drug

combinations. *Cancer Research*, 81(17), 4603–4617. <https://doi.org/10.1158/0008-5472.can-21-1208>

### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Figure S1:** Transition graphs of inferred models, including neutral.

**How to cite this article:** Paulevé, L., & Gaucherel, C. (2025). Inference of ecological networks and possibilistic dynamics based on Boolean networks from observations and prior knowledge. *Methods in Ecology and Evolution*, 00, 1–17. <https://doi.org/10.1111/2041-210X.70090>