

Is a community state reachable, and why?

Mathieu de Goër de Herve^{1,a,*}, Colin Thomas^{2,1,b}, Maximilien Cosme^{1,c},
Philip H. Warren^{3,d} and Cédric Gaucherel^{1,e}

January 14, 2022

¹ AMAP - INRAE, CIRAD, CNRS, IRD, Université de Montpellier, Montpellier, France

² IBISC, Univ. Évry, Univ. Paris-Saclay, 91020 Évry-Courcouronnes, France

³ School of Biosciences, University of Sheffield, Sheffield S10 2TN, UK

^amathieu.de-goer@protonmail.com, ^bcolin.thomas@univ-evry.fr, ^cmaximilien.cosme@cirad.fr,
^dp.warren@sheffield.ac.uk, ^ecedric.gaucherel@inrae.fr

Statement of authorship: MGH led the project with initial inputs from CT and CG. MGH performed the analyses and wrote the first draft from the manuscript. All authors contributed substantially to idea development and revisions of the manuscript.

Data accessibility statement: If the manuscript gets accepted, the data supporting the results will be archived in a public repository and the data DOI will be included at the end of the article.

Keywords: Assembly, Rule-based model, Qualitative model, Microcosms, Reachability, Assembly Pathways, Discrete-event model, Community collapse.

Type of article: Letter

Document statistics:

144 words in the abstract

4981 words in the main text

35 references

5 figures, 1 table, 0 text boxes

* Corresponding author: Mathieu de Goër de Herve (mathieu.de-goer@protonmail.com)

Abstract

Deterministic models have difficulties to take into account stochasticity during community assembly. As a tool to circumvent this problem, we present a qualitative discrete-event model, where consequences of interspecific interactions are described as rules. This model provides a map of all possible future dynamics for a given system, which allows to exhaustively describe the possible pathways during an assembly process. Such a description does not rely on species traits details and is insensitive to stochastic effects. This allows to show that subsets of species are sometimes impossible to reach starting from larger sets of species, and therefore to question the reachability of community states during the system's dynamics. Applying the model to an experimental dataset studying the collapse of protist communities, we obtain a very good theory-experiment agreement. We finally discuss what the notion of reachability can bring to community assembly.

1 Introduction

Ecological communities are complex and subject to stochastic behaviour, which makes them very difficult to model. A common approach is to use systems of differential equations and assume that stochastic behaviour manifests as fluctuations around the equilibrium (Lotka, 1925; Volterra, 1926; Chesson, 2000). However, real ecological communities commonly lose and gain species, leading to out-of-equilibrium behaviour in which stochasticity and historical contingency play important roles (Fukami, 2015; DeMalach *et al.*, 2021). Deterministic models can predict possible stable states, but their limitations far away from equilibrium hinder their ability to describe shifts between these states.

These difficulties are particularly striking for studying community assembly, that is how communities form through successive invasions and species loss. The order of invasion sequences as well as their timing and the number of invaders can lead to very different final communities (Drake, 1991) ; all of these parameters are random in nature and make the study and modelling of assembly process particularly challenging (Drake *et al.*, 2012; Song *et al.*, 2021). Modelling efforts in this field led to draw assembly graphs, whose nodes are the system’s equilibrium points and edges are transitions between them caused by invasions (Capitán *et al.*, 2009; Serván & Allesina, 2021). However, issues over the computation of assembly graphs, as well as more conceptual questions such as the ecological importance of transient states, remain open to debate (Fukami & Nakajima, 2011). Moreover, these models still lack experimental validation.

This paper aims to provide a family of models able to describe the long-term dynamics of a community while taking into account its possible stochasticity. Although a quantitative model able to provide probabilistic predictions on the community composition at a future date might sound desirable, the practical feasibility of such a model is unclear due to the large computational power required. Furthermore, the large number of experimental replicates necessary to measure probability distributions is usually unworkable, making such a model extremely difficult to validate. We therefore choose to model communities qualitatively, considering only the presence or absence of species. Although not so common, such models have been present for a long time in community ecology (Luh & Pimm, 1993), to address questions such as the cascading extinctions caused by removing particular species from a trophic network (Srinivasan *et al.*, 2007) or the effects of ecosystem engineering on community assembly (Yeakel *et al.*, 2020).

Since species can appear or disappear, the community composition at a given time is considered as a subset of a possible global species pool, corresponding for example to species that are regionally present (Fukami, 2015). Based on the interaction network of the pool, one can write down a set of logical rules describing how these interactions can affect the community composition and which events (e.g. extinctions or invasions) they can cause (Gaucherel & Pommereau, 2019; Veloz, 2020; Mao *et al.*, 2021). The subset of species present at a given time within the system then determines which of these rules may or not be applied, and which composition changes they would cause to the system - and therefore, which new species sets can follow. Applying this process to each possible species combination allows to draw an exhaustive map of successive possible

species subsets. The model does not provide information on whether a future subset is more or less likely than an other one: it discriminates between which are *possible* and which are not - or, put differently, it tells which subsets may be reached starting from a given one as well as the possible pathways, regardless from traits, abundances or stochastic fluctuations.

Such models have already been successfully used in biology to describe reaction networks, which present similar challenges of complexity and stochasticity (Chabrier & Fages, 2003; Wang *et al.*, 2012). Applying them to ecological systems seems promising and possibly very powerful, but it requires experimental validation. Datasets suitable to do so are rare, as it requires long-term tracking of a community that gains and/or loses species as well as enough replicates to explore possibly branching pathways in community composition. We test this new approach on previously-unpublished data, taken from an experiment on protist species (Weatherby *et al.*, 1998). This experiment is unique, as far as we know, in being the only replicated analysis which systematically explores the fate of all possible communities that can be built from a pool of, in this case, six species (i.e. a powerset of $2^6 - 1 = 63$ communities in total). The experiment was originally done in order to find all the persistent sets of species in the powerset (Weatherby *et al.*, 1998), to first examine the invasibility of these sets (Law *et al.*, 2000), and then use this information to map the possible assembly pathways for the entire system (Warren *et al.*, 2003). Data were recorded from monthly censuses of the replicate communities, documenting the changes in species composition as communities moved along their pathways to the final persistent communities. These time series contain useful information about the order in which species were lost, including random variation across replicates, on which the power of rule-based models can be assessed. We use this experiment to test and validate our proposed approach, and to demonstrate how it can improve our understanding of long-term community dynamics and assembly.

2 Materials and methods

2.1 The experiment

We summarize here the experiment fully described in (Weatherby *et al.*, 1998), from which the data we used are derived. Experimental communities were constructed from combinations of six species of protists plus a mixed bacterial flora. Protist species were chosen to exhibit a range of sizes and trophic strategies, and on the basis that they were all able to persist when grown under the same environmental and nutrient medium conditions in the laboratory. The trophic relationships of these species based on prior information of their feeding habits allow drawing the trophic network under study (Fig. 1). Before the experiments started all species were kept in laboratory stock cultures using a nutrient regime similar to that used in the experiments, while predator populations *Amoeba proteus* and *Euplotes patella* were fed on *Tetrahymena pyriformis*. Hereafter, we will refer to each species by the initial letter of the genus name: *A*, *B*, *P*, *E*, *C* or *T* in the community notations (the ordering of species names is according to

their size).

Each of the 63 combinations of protist species was replicated six times over the whole experiment (378 microcosms in total). For each combination, approximately 100 individuals of each species were introduced in the microcosms, with a mixture of bacteria. Microcosms were sampled on day 17, and then every 28 days until specified stopping criteria were reached. These were either: (i) species composition was unchanged for three successive samples; (ii) only one species known to be able to persist alone remained; (iii) only one species known not to be able to persist alone remained; (iv) all species were extinct. In the original experiment, most (93%) microcosms had met one of these criteria by 101 days, and the composition at this point was taken as the ‘final’ state for the purposes of the analyses in Weatherby (1998). However, here we make use of the full time series information to examine the pathways of the replicate systems to inform and validate the model. These time series data consist of information from monthly samples of every microcosm, which involved systematically scanning the entire microcosm, in situ, under 20-30 times magnification and recording the presence/absence of each species, along with an estimate of abundances. In this article, we only use the presence/absence data.

While the majority of the experiments were reproducible, some were not. An example of the results for such an experiment are pathways starting with species A,P,T (Table 1). The lack of reproducibility between replicates is here striking: neither the final result, nor the pathways, nor the timing, are consistent. In the complete data set, at least 32 out of the 63 experiments displayed multiple different final compositions and/or pathways. Such variability in the outcomes of experiments is not surprising for ecological systems, even in well-controlled conditions, and better understanding this variability is an active research topic (Goldford *et al.*, 2018). Here, we provide an appropriate model to describe such behaviour, and we will illustrate our approach with a toy-model using this APT subsystem.

Start	day 14	42	70	98	126	154	182	210	238	266
APT	APT	AT	AT	AT	AT	A	–			
APT	APT	AT	AT	T						
APT	APT	APT	AP	A	A	–				
APT	APT	APT	APT	AP	AP	AP	AP	A	A	–
APT	APT	APT	APT	AT	AT	AT	A	–		
APT	APT	A	A	A	–					

Table 1: Time evolution of the community composition for each replicate of the APT experiment. “–” corresponds to no species left (and the corresponding replicate was then terminated). Replicate 2 was terminated once T was left alone as it is known to survive arbitrarily long by feeding on bacteria.

2.2 A toy-model for our approach

Similar to the approaches used in (Luh & Pimm, 1993; Law & Morton, 1996; Capitán *et al.*, 2009; Campbell *et al.*, 2011), we describe the system without considering population abundances. We define the community state as the set of species present in the system. For example, if both A and P are present (and thus B,E,C,T absent), the state of the system will be denoted $\{A, P\}$. The interactions between species are then translated as *rules*, describing how each species *could* affect another one in a way that changes the community composition. A rule describes an event that can happen to the community and change its composition. It contains the conditions necessary for this event to happen and its consequences on community composition, written as *conditions* \rightarrow *consequences*. For example, trophic relation between populations A and P (predation of A on P) can lead to two events: A can drive P to extinction, and A can go extinct if P is absent (assuming here that no other prey species is present). Denoting by + (−) the species that are present (absent), the first rule can be written “ $A+, P+ \rightarrow P-$ ”: if A and P are both present (+), then P can become absent (−). For the sake of concision, this rule is rewritten simply as $A+ \rightarrow P-$. Indeed, if P is already absent from the system, applying the rule will not change anything (a rigorous justification for this writing choice is found in (Gaucherel & Pommereau, 2019)). The rule associated to the second event (starvation of A) is then “ $A+, P- \rightarrow A-$ ”, simplified into $P- \rightarrow A-$. We do not provide probabilities for the events; we are interested in exhaustively describing what *can* happen, and whether these changes happen or not will depend on the details of traits, abundances, or contexts.

Once the rules describing what can happen to the community are formalized, it is possible to identify for each possible state which rules are enabled and which state they lead to. This allows computation of the graph made up of all states and all transitions between them, hereafter called *state space*. This graph is similar to the “assembly graph” representation which is common in community assembly (Law & Morton, 1996; Capitán *et al.*, 2009; Serván & Allesina, 2021), except that our model does not consider which states might or might not be stable: we draw *all* possible states and *all* transitions between them.

We apply this model to the experiment containing only species A, P and T (Fig. 2). Comparing the predicted pathways with the results of table 1 displays a remarkable agreement: all transitions between states observed in the experiments are predicted on the graph, and all transitions predicted on the graph are observed in the experiments. This model, despite its simplicity, describes successfully the possible transitions and pathways between the possible community states, based solely on the nature of inter-specific interactions. Interestingly, the state space is not a connected graph: this implies that state $\{P\}$ can never be reached when all three species are joined together, regardless the initial species abundances of the system or stochasticity. This (un)reachability property is one of the central conclusions of this work and will be discussed in more details below. The transition from $\{P, T\}$ to $\{P\}$ is not observed in table 1 as $\{P, T\}$ cannot be reached from $\{A, P, T\}$; it has however been observed in other sets of experiments.

2.3 Model for the full set of experiments

To model the complete set of experiments, we first formalize as rules the possible consequences of interactions between species at the scale of the community. Here, species can only go extinct, and defining rules therefore corresponds to determining the conditions in which they can disappear. We identified three types of interactions in this species community:

- **Predation.** The first consequence of such a relationship is that the predator (population) may drive its prey (populations) to extinction. The second one is that the predator can go extinct if it has no more prey. The predation of a species Pr upon prey N will therefore be written as two rules: $Pr+ \rightarrow N-$ and $N- \rightarrow Pr-$. If it has two prey N_1 and N_2 , we will assume that the predator can disappear only if none of its prey is there, and there will be three rules instead of four: $Pr+ \rightarrow N_1-$, $Pr+ \rightarrow N_2-$ and $N_1-, N_2- \rightarrow Pr-$.
- The system described in (Weatherby *et al.*, 1998) displayed a second kind of predation, where the prey population can be unable to sustain the population of the predator for a number of generations (Fig. 1). This means that this kind of prey should not be considered when determining whether the predator can starve or not; however, the predator might still be able to drive them to extinction. We chose to describe it as a secondary predation, meaning that the predator will go towards its “preferred” prey first. Therefore, it can drive the “secondary” ones to extinction if the preferred ones are absent. With a secondary predation interaction upon N_s to the system, we need to add the rule $Pr+, N_1-, N_2- \rightarrow N_s-$.
- **Competition.** Its consequences are simple: one of the competitors can be excluded. The competition between two species S_1 and S_2 is therefore written as two rules: $S_1+ \rightarrow S_2-$ and $S_2+ \rightarrow S_1-$. If the competition is strongly unbalanced and one of the species is the dominant competitor in all contexts, only one of these rules will be used.

At this stage, a first attempt to model the community of the experiment set would be to derive the rules directly from the trophic network drawn in (Weatherby *et al.*, 1998) (Fig. 1). However, this attempt gave disappointing results, because in many experiments starting with several basal species one of them went extinct even in the absence of predators. It suggests that competition between species must be taken into account in the model. We therefore chose to infer the whole interaction network from the outcomes of the two-species experiments (Fig. 3). If a prey population is able to sustain its predator’s population for a number of generations, the predator should never disappear during the experiments involving only the predator and its prey as long as the prey is present. If not, the predator might go extinct before the prey in some replicates. Similarly, the competition between basal species can be deduced from the result of pairwise experiments. We also observed in the data that species C sometimes disappeared alone, for unclear reasons, and therefore added the rule $C+ \rightarrow C-$. The rules that subsequently became redundant were then removed; for example $B+ \rightarrow C-$

(predation of B upon C) was removed because $C+ \rightarrow C-$ already allows C to disappear in every situation where $B+ \rightarrow C-$ can be applied. Although both rules have different meanings, keeping only $C+ \rightarrow C-$ leads to the same computation result. Similarly, E has been found to have no species on which it can reliably feed upon on the long term and can therefore go extinct in any situation, and the rule corresponding to secondary predation of A upon E was removed because of its redundancy with $E+ \rightarrow E-$ (Fig. 3).

3 Results

3.1 Complete state space and model validation

Once the community interaction network and its associated ruleset are defined (Fig. 3), the state space is computed from these rules (Fig. 4). In the experiments, species disappear one by one and never reappear. There are therefore $\sum_{k=1}^6 k \binom{6}{k} = 6 \times 2^{6-1} = 192$ theoretically possible transitions between the $2^6 = 64$ states (six transitions starting from the state with six species, five transitions starting from each of the $\binom{6}{5}$ states with five species, etc). Out of these 192 transitions, 135 are predicted by the model, among which 71 (53%) are observed. No unpredicted transition is observed, which is a necessary condition for validating the model (Saadatpour & Albert, 2013).

It is expected that some of the predicted transitions are never observed, for two reasons. First, transitions with a very low probability may be unobserved due to the finite number of replicates. In particular, the states with many initial species always collapse towards states with less species; transitions starting from the latter are therefore observed more often. For example, the state $\{A, B\}$ was observed 77 times. Starting from this state, the transition to $\{B\}$ was observed 73 times and the transition to $\{A\}$ only three times. Such an infrequent transition would probably not have been observed if $\{A, B\}$ had been observed only six times (like $\{A, B, P, E, T\}$ for example). Secondly, some transitions were too fast to be visible within the data. For example, all replicates starting from all 6 species had only four species left at the first census, and the transitions to 5-species states were therefore not observed even if they must have happened (assuming that no two species disappeared at the same time). Finally, the model is parameter-free and ignores details on abundances and species traits: transitions are predicted if they can occur in *some* conditions, while these conditions might never have been reached during the experiments. The model over-predicts possible pathways on purpose, in order to *exhaustively* predict what may happen.

This choice of a broader (qualitative and discrete) prediction does not prevent providing information on the system’s dynamics. However, it becomes crucial to know to what extent we over-predict and to ensure that all observed transitions were not predicted by chance due to a too broad prediction. For this purpose, we used a null-hypothesis test based on a neutral (random) model. As the transitions from two to one or one to zero species are used to calibrate the model, we considered only transitions

starting from three or more species to validate it.

In the experiment, $\sum_{k=3}^6 k \binom{6}{k} = 156$ of these transitions are possible. In the complete community model, 111 of these 156 possible transitions are predicted, among which 51 (46%) are observed in the experiment. If 111 transitions were randomly predicted out of the 156 (corresponding to the case where the model would provide no information), the probability to observe only predicted transitions would therefore be $\binom{111}{51} / \binom{156}{51} = 3 \times 10^{-10}$ (also equal to the probability that 51 transitions drawn randomly out of 156 are all predicted). This allows us to be confident in the capacity of our model to indeed predict all possible pathways of the community.

3.2 Are there stable states?

One could hope to use this model to predict stable states from the community, looking for the states from which no transition is possible (called “deadlocks” (Gaucherel & Pommereau, 2019)). However, in our complete community model there is only one such state containing more than a single species ($\{B, P\}$, in purple on Fig. 3). While it might be surprising at first sight, this is indeed what is observed within the data (see supplementary S1). Some states collapsed quickly and were never observed at two successive census while others usually lasted for dozens of generations, but all states but $\{B, P\}$ containing two or more species decayed at some point. The reasons for this behaviour are unclear, possibly a lack of control upon experimental conditions or evolutionary effects during these long-term experiments. However, as our model does not depend on exact parameters and traits, it should be robust to these effects and indeed appears to be. It also illustrates that this discrete-event model does not make any assumption on timescales of processes and events, nor on how long the system will stay in a given state. The deadlocks in this model are in fact highly different in nature from the stable states commonly studied in community ecology. Indeed, stable states usually correspond to an equilibrium between processes and therefore to a restricted part of the phase space. As our model should remain valid regardless from abundances and parameters, deadlocks rather correspond to an absence of processes allowing to leave the state.

3.3 Reachability of the states

Strikingly, the state space is disconnected (Fig. 4). This implies that all states cannot be reached from the community containing all species. This includes stable states: $\{B, P\}$ is highly stable but cannot be reached from $\{A, B, P, E, C, T\}$. As the model is parameter-free, this impossibility to go from a given state to another should remain true for any parameter set, as long as the topology of the system – corresponding to the set of rules – remains unchanged (see Discussion). This result should also be robust to stochasticity: parameter changes and stochasticity would affect which exact pathway is taken by the system, but this pathway should stay within the set of the computed possible transitions.

This non-reachability property can be understood by considering species A and P, two key populations of this community. Species A is an apex predator while P, a bacterivorous species, is able to sustain itself for an arbitrarily long duration. The only way A can go extinct is therefore its starvation in the absence of P, and P will thus necessarily disappear before A. States where P is present but A is absent are therefore unreachable from states where both are present. On the other hand, no species other than A is able to drive P to extinction. Hence, a state without P cannot be reached from a state with P but without A. As simple as this explanation is, such a property is not visible without drawing the community’s state space with our model.

Computer sciences provide a range of methods for automatically analysing reachability. It is possible to group states into “structures” sharing the same topological properties in the state space (e.g. same possible final outcomes (Bérengruier *et al.*, 2013; Gaucherel & Pommereau, 2019)). In the state space (Fig. 4), states leading to the same endpoints have been plotted with the same color. They can be “merged” to provide a convenient state space which summarizes the overall dynamics and displays all states belonging to the same structure as a single super-node (Fig. 3 inset). This highlights key transitions having causal consequences in the overall dynamics and causing shifts from a structure to another one (Gaucherel *et al.*, 2020). For example, starting from $\{A, B, E, C, T\}$ (dark blue, second line on figure 3), losing species A implies that the sole possible final state of the community is $\{B\}$, regardless of future dynamics details (as state $\{B, E, C, T\}$ belongs to the yellow structure). Conversely, losing species C or E would have fewer consequences, as it has no impact on the long-term reachable states: both states $\{A, B, E, T\}$ and $\{A, B, C, T\}$ belong to the same topological structure as $\{A, B, E, C, T\}$. This method is particularly useful for dealing with large species communities or ecosystems, where the state space is impossible to plot directly because the number of states, 2^n , increases exponentially with the number of variables n in the system (Cosme *et al.*, 2022).

3.4 Predicting the consequences of invasions

Although our model mainly describes the possible collapses of communities, it can also provide insights on invasion consequences, by considering invasion as a one-time event that may happen starting from a given state. Partitioning the state space into its topological structures is especially useful here. Depending on which species invade, invasions can be sorted into three categories (Fig. 5):

- The consequence of an invasion can be weak, when the post-invasion state belongs to the same structure than the one containing the initial state. In this case, the invaded state might collapse back towards the initial state, or follow the same long-term future than the one of the initial state. For example, invasion of state $\{B, T\}$ by species C pushes the community into $\{B, C, T\}$, which can collapse back to $\{B, T\}$ and finally reach the same final endpoint $\{B\}$.
- The consequence of an invasion can be an irreversible shift when it leads to a state that cannot come back to the initial state nor reach similar endpoints. For

example, invasion of $\{B, T\}$ by P leads to state $\{B, P, T\}$, which always ultimately leads to $\{B, P\}$.

- Finally, the consequences can be indeterminate if the invaded state can either come back to the initial state or collapse to states belonging to other structures. For example, invasion of state $\{B, T\}$ by A leads to $\{A, B, T\}$ which can collapse back to $\{B, T\}$, but also towards $\{A, T\}$, leading to an irreversible shift for the future of the community.

These three invasions of state $\{B, T\}$ have been experimentally proven possible in (Warren *et al.*, 2003), although the subsequent long-term dynamics was not studied.

4 Discussion

The study of long-term dynamics of communities and community assembly requires models that avoid sinking into unmanageable complexity, and able of dealing with the unavoidable stochasticity at this scale. Here, we proposed a simple qualitative framework based on a rule-based and discrete-event formalism. The state of the studied community is defined by species which are present in it, then associated with possible changes in community composition caused by interactions within the system. Such models provide an exhaustive map (the state space) of possible pathways for the community or ecosystem (Gaucherel & Pommereau, 2019; Cosme *et al.*, 2022). Although this modelling framework might seem rudimentary at first sight, we applied it to an experiment studying the long-term dynamics of protist communities (Weatherby *et al.*, 1998) and demonstrated its good capacity to describe the experiments, with an excellent agreement between model and data. The model is not predicting the exact final state that a specific community will end up in, but it allows to discriminate between which states are reachable and which are not. Focusing on reachability allows to better understand the dynamics of an ecological system while taking stochasticity into account, and without relying on unavailable information and parameters. This model also automatically sorts the system states and pathways depending on their reachable endpoints and provides a synthesized map of all possible fates. Finally, it also allows some insights into the possible consequences of a given invasion event, based on the options for the system to come back to its initial state and the possible impact on the system’s long-term changes.

4.1 Consequences for community assembly

Our approach appears useful for experiments in community assembly. The “top-down” experiments, in which all species are put together before letting the system evolve (Goldford *et al.*, 2018; Bittleston *et al.*, 2020), are often used as an easier-to-perform alternative to “bottom-up” experiments, in which species are added one by one into the system. Hence, a better understanding of the relationship between both approaches would have

a substantial interest (Serván & Allesina, 2021). Studying reachability provides insights about situations where both directions are not equivalent. For example, the (highly stable) state $\{B, P\}$ cannot be reached from the initial state $\{A, B, P, E, C, T\}$, while an other set of experiments with the same species showed that it could be built through invasion of state $\{B\}$ by species P and reversely (Warren *et al.*, 2003). This is a clear demonstration of a case where top-down and bottom-up assemblies differ. Also, we would like to point out that assembly models often focus on invasions as the main source of uncertainty while the collapse of unstable states following invasions is assumed deterministic; here we show that this collapse is also far from trivial and should not be neglected.

Community assembly is sometimes studied using “assembly graphs” (Law & Morton, 1993; Capitán *et al.*, 2009; Serván & Allesina, 2021). This approach considers only stable states and transitions between them, howether it has been demonstrated that stable endpoints of community assembly are not necessarily representative of transient states and that transient states have no reason to be short-lived enough to be negligible (Fukami & Nakajima, 2011). In the set of experiments studied here, only one state containing more than a single species was observed as truly persistent, all the other states losing species at some point (see Results). Like in (Fukami & Nakajima, 2011), it seems that a large part of the interesting dynamics lies outside from stable states. It should however be possible to combine both approaches, by using differential equations to identify which states can be stable and our rule-based model to identify possible transitions.

4.2 A qualitative model

Using a qualitative framework comes with many advantages despite removing the opportunity to provide precise predictions on a given community’s future. Firstly, on the side of experimental and empirical works, it allows writing meaningful models even in the absence of exhaustive measurements of all parameters, often tedious if not impracticable. For example, in the case of the Weatherby 1998 experiments, the detailed composition of the bacteria on which protists are fed was unknown, while it has been recently demonstrated that basal resources plays a crucial role in assembly (Goldford *et al.*, 2018; Dal Bello *et al.*, 2021). Here, while this composition may play a role in the detailed dynamics of the microcosms, it should not change what *can* happen: it may affect what *will* happen, i.e. which exact pathways will be followed within the possible ones. The computed state space provides here a canvas showing what is possible and what is not, regardless of parameter values. For this reason, the model could also be useful for experimental design, as it can provide with a few modelling efforts an *a priori* description of possible outcomes without requiring many preliminary measurements. Secondly, on the theoretical side, a central advantage of this parameter-free model is that its predictions are highly robust. Indeed, they should be valid for any set of parameters as long as the topology of the ecological network remains identical (same interactions). It should also be robust to most higher order interactions, as long as they belong to trait-mediated indirect interactions (Golubski *et al.*, 2016; Levine *et al.*,

2017): since our model is designed to describe all possible transitions regardless from trait values, higher-order interactions that affect traits should not add new transitions. Instead of exploring various sets of parameters, the model encompasses all possible values in one single result and provides a “universal” canvas within which the detailed dynamics occur. This can be compared to a recent paper (Yeakel *et al.*, 2020), in which the authors randomly draw the changes in a species community based on logical rules. While their approach would correspond to performing a weighted random walk in the state space, our method provides a tool to directly study the graph within which the dynamics occur as a whole, which provides complementary information.

Lastly, this parameter-free framework and these models are fast to compute: a few seconds only are needed for the graphs presented here, and a few tens of minutes for huge models containing around a hundred variables without using a supercomputer (Cosme *et al.*, 2022). The work realized on biochemical networks even demonstrated the ability to handle networks of more than 500 variables (each species corresponds to one variable in our case) (Chabrier & Fages, 2003). Visualizing the corresponding state spaces becomes rapidly impossible as its size increases exponentially with the number of variables, but many tools do exist in computer sciences for analyzing such models. For example, methods derived from model checking (Clarke *et al.*, 2018) could be implemented to efficiently compute the reachable states and question various dynamical properties.

4.3 Conclusion

We demonstrated that a simple and qualitative model, based on a translation of the interaction network as a set of logical rules, could effectively provide information on the long-term dynamics of an ecological community. Choosing a qualitative model and focusing on the essence of interactions rather than on their details provides a global view that is unaffected by the stochastic variations inherent to ecological systems. We hope and believe that this approach can provide new insights on the long-term dynamics of communities and community assembly that would be unattainable with quantitative models.

5 Acknowledgments

We warmly thank Franck Pommereau, Sonia Kéfi and Richard Law for their suggestions and advice on the study and the manuscript, and Anita Weatherby for help with, and discussions about, the experimental data. We also thank the SESASA project (Leap-Agri European call) for its financial support of this work. The original experimental work from which the data derives was funded under the UK Natural Environment Research Council grant GST/91/870.

References

- Bittleston, L. S., Gralka, M., Leventhal, G. E., Mizrahi, I. & Cordero, O. X. (2020). Context-dependent dynamics lead to the assembly of functionally distinct microbial communities. *Nature Communications*, 11, 1440.
- Bérengruier, D., Chaouiya, C., Monteiro, P. T., Naldi, A., Remy, E., Thieffry, D. & Tichit, L. (2013). Dynamical modeling and analysis of large cellular regulatory networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23, 025114.
- Campbell, C., Yang, S., Albert, R. & Shea, K. (2011). A network model for plant–pollinator community assembly. *Proceedings of the National Academy of Sciences*, 108, 197–202.
- Capitán, J. A., Cuesta, J. A. & Bascompte, J. (2009). Statistical Mechanics of Ecosystem Assembly. *Physical Review Letters*, 103, 168101.
- Chabrier, N. & Fages, F. (2003). Symbolic model checking of biochemical networks. In: *Computational Methods in Systems Biology* (ed. Priami, C.). Springer Berlin Heidelberg, Berlin, Heidelberg.
- Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, 31, 343–366.
- Clarke, E. M. J., Grumberg, O., Kroening, D., Peled, D. & Veith, H. (2018). *Model Checking, Second Edition*. MIT Press.
- Cosme, M., Hély, C., Pommereau, F., Pasquariello, P., Tiberi, C., Treydte, A. & Gaucherel, C. (2022). Qualitative modeling for bridging expert-knowledge and social-ecological dynamics of an east african savanna. *Land*, 11.
- Dal Bello, M., Lee, H., Goyal, A. & Gore, J. (2021). Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism. *Nature Ecology & Evolution*, 5, 1424–1434.
- DeMalach, N., Shnerb, N. & Fukami, T. (2021). Alternative states in plant communities driven by a life-history trade-off and demographic stochasticity. *The American Naturalist*, 198, E27–E36.
- Drake, J. A. (1991). Community-Assembly Mechanics and the Structure of an Experimental Species Ensemble. *The American Naturalist*, 137, 1–26.
- Drake, J. A., Staelens, P. & Wieczynski, D. (2012). Assembly processes. In: *Encyclopedia of Theoretical Ecology* (eds. Hastings, A. & Gross, L.). University of California Press.
- Fukami, T. (2015). Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. *Annual Review of Ecology, Evolution, and Systematics*, 46, 1–23.
- Fukami, T. & Nakajima, M. (2011). Community assembly: alternative stable states or alternative transient states? *Ecology Letters*, 14, 973–984.

- Gaucherel, C. & Pommereau, F. (2019). Using discrete systems to exhaustively characterize the dynamics of an integrated ecosystem. *Methods in Ecology and Evolution*, 10, 1615–1627.
- Gaucherel, C., Pommereau, F. & Hély, C. (2020). Understanding Ecosystem Complexity via Application of a Process-Based State Space rather than a Potential Surface. *Complexity*, 2020, 1–14.
- Goldford, J. E., Lu, N., Bajić, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., Segrè, D., Mehta, P. & Sanchez, A. (2018). Emergent simplicity in microbial community assembly. *Science*, 361, 469–474.
- Golubski, A. J., Westlund, E. E., Vandermeer, J. & Pascual, M. (2016). Ecological Networks over the Edge: Hypergraph Trait-Mediated Indirect Interaction (TMII) Structure. *Trends in Ecology & Evolution*, 31, 344–354.
- Law, R. & Morton, R. D. (1993). Alternative Permanent States of Ecological Communities. *Ecology*, 74, 1347–1361.
- Law, R. & Morton, R. D. (1996). Permanence and the Assembly of Ecological Communities. *Ecology*, 77, 762–775.
- Law, R., Weatherby, A. J. & Warren, P. H. (2000). On the invasibility of persistent protist communities. *Oikos*, 88, 319–326.
- Levine, J. M., Bascompte, J., Adler, P. B. & Allesina, S. (2017). Beyond pairwise mechanisms of species coexistence in complex communities. *Nature*, 546, 56–64.
- Lotka, A. (1925). *Elements of Physical Biology*. Williams & Wilkins, Baltimore, MD.
- Luh, H.-K. & Pimm, S. L. (1993). The Assembly of Ecological Communities: A Minimalist Approach. *The Journal of Animal Ecology*, 62, 749.
- Mao, Z., Centanni, J., Pommereau, F., Stokes, A. & Gaucherel, C. (2021). Maintaining biodiversity promotes the multifunctionality of social-ecological systems: holistic modelling of a mountain system. *Ecosystem Services*, 47, 101220.
- Saadatpour, A. & Albert, R. (2013). Boolean modeling of biological regulatory networks: A methodology tutorial. *Methods*, 62, 3–12.
- Serván, C. A. & Allesina, S. (2021). Tractable models of ecological assembly. *Ecology Letters*, 24, 1029–1037.
- Song, C., Fukami, T. & Saavedra, S. (2021). Untangling the complexity of priority effects in multispecies communities. *Ecology Letters*, 24, 2301–2313.
- Srinivasan, U. T., Dunne, J. A., Harte, J. & Martinez, N. D. (2007). Response of complex food webs to realistic extinction sequences. *Ecology*, 88, 671–682.
- Veloz, T. (2020). The Complexity–Stability Debate, Chemical Organization Theory, and the Identification of Non-classical Structures in Ecology. *Foundations of Science*, 25, 259–273.

- Volterra, V. (1926). Fluctuations in the Abundance of a Species considered Mathematically 1. *Nature*, 118.
- Wang, R.-S., Saadatpour, A. & Albert, R. (2012). Boolean modeling in systems biology: an overview of methodology and applications. *Phys. Biol.*, 15.
- Warren, P. H., Law, R. & Weatherby, A. J. (2003). Mapping the Assembly of Protist Communities in Microcosms. *Ecology*, 84, 1001–1011.
- 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, 554–566.
- Yeakel, J. D., Pires, M. M., de Aguiar, M. A. M., O'Donnell, J. L., Guimarães, P. R., Gravel, D. & Gross, T. (2020). Diverse interactions and ecosystem engineering can stabilize community assembly. *Nature Communications*, 11, 3307.

List of Figures

1	Trophic network of the experimental system with species sizes indicated, as displayed in (Law <i>et al.</i> , 2000). Solid lines are feeding links capable of sustaining a species for a number of generations; broken lines are feeding links unable to do so.	18
2	Illustration of our model with a subset of the species present in the experiments: A, P and T. a) Interaction network of the species. A is a predator for P and T (blue arrows), but the population of T alone cannot sustain the population of A for many generations (blue dashed). P and T compete together (red arrow), but P is always the dominant competitor. This network is inferred from the experiments involving pairs of species (see section 3). b) List of the rules used to describe the system (see main text). c) <i>State space</i> of the system. The rules shifting from one state to another one are labelled on edges. As the system is isolated, invasions are impossible and the graph is acyclic. States that have the same possible final outcome (called a deadlock) are displayed with the same color; states with several possible outcomes are drawn as diamonds while square states have only one possible final outcome.	19
3	Trophic network deduced from the pairwise experiments, with the associated ruleset (to be compared to Fig. 1). Blue arrows correspond to predation interaction (dashed if the prey cannot sustain the predator’s population alone), red arrows to competition interactions. Species with a magenta circle are observed to be able to survive alone, by feeding on bacteria. We chose to describe the predation of A upon P as preferential, as A can sustain itself only on P: we assume that it will not drive other species to extinction as long as P is present.	20
4	Complete graph of the 64 possible states for the system, with the transitions predicted by the model (defined in Fig. 3). Black (grey) arrows correspond to observed (unobserved) transitions in the experiments. States drawn with the same color have the same possible final states. Square states have only one possible final outcome; diamonds several ones. <u>Inset</u> : schematic (“merged”) graph of the complete state space, where all states with the same possible outcomes are merged into one super-node. Edges in this merged graph indicate that there exist transitions predicted between <i>some</i> of the states within super-nodes. Labels indicate which species are present in <i>all</i> the states of a given super-node.	21
5	Subset of the state space (Fig. 4) corresponding to the invasion of state $\{B, T\}$ by species C, A or P (blue dotted arrows). Starting from $\{B, T\}$, invasion by C has a weak impact, invasion by P has a large and irreversible impact and invasion by A has an indeterminate effect (see text).	22

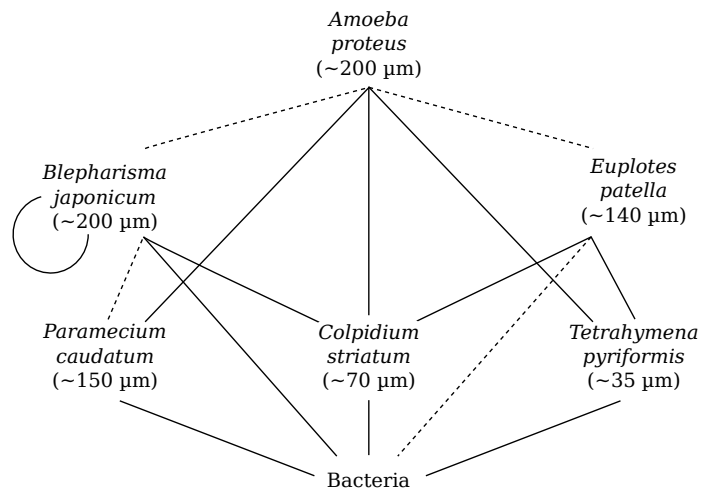


Figure 1: Trophic network of the experimental system with species sizes indicated, as displayed in (Law *et al.*, 2000). Solid lines are feeding links capable of sustaining a species for a number of generations; broken lines are feeding links unable to do so.

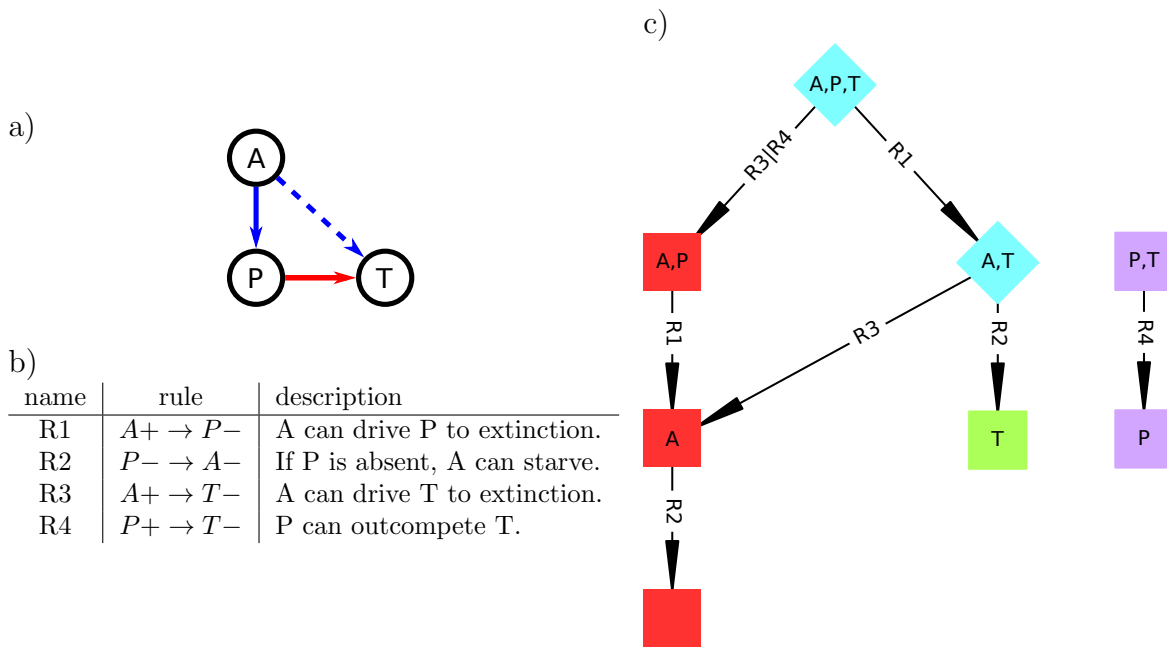


Figure 2: Illustration of our model with a subset of the species present in the experiments: A, P and T. a) Interaction network of the species. A is a predator for P and T (blue arrows), but the population of T alone cannot sustain the population of A for many generations (blue dashed). P and T compete together (red arrow), but P is always the dominant competitor. This network is inferred from the experiments involving pairs of species (see section 3). b) List of the rules used to describe the system (see main text). c) *State space* of the system. The rules shifting from one state to another one are labelled on edges. As the system is isolated, invasions are impossible and the graph is acyclic. States that have the same possible final outcome (called a deadlock) are displayed with the same color; states with several possible outcomes are drawn as diamonds while square states have only one possible final outcome.

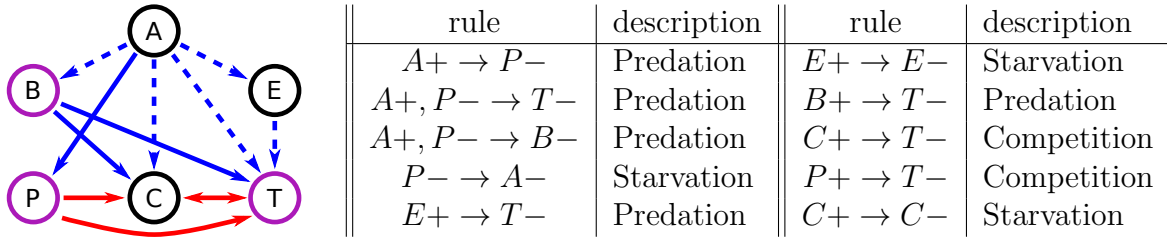


Figure 3: Trophic network deduced from the pairwise experiments, with the associated ruleset (to be compared to Fig. 1). Blue arrows correspond to predation interaction (dashed if the prey cannot sustain the predator’s population alone), red arrows to competition interactions. Species with a magenta circle are observed to be able to survive alone, by feeding on bacteria. We chose to describe the predation of A upon P as preferential, as A can sustain itself only on P: we assume that it will not drive other species to extinction as long as P is present.

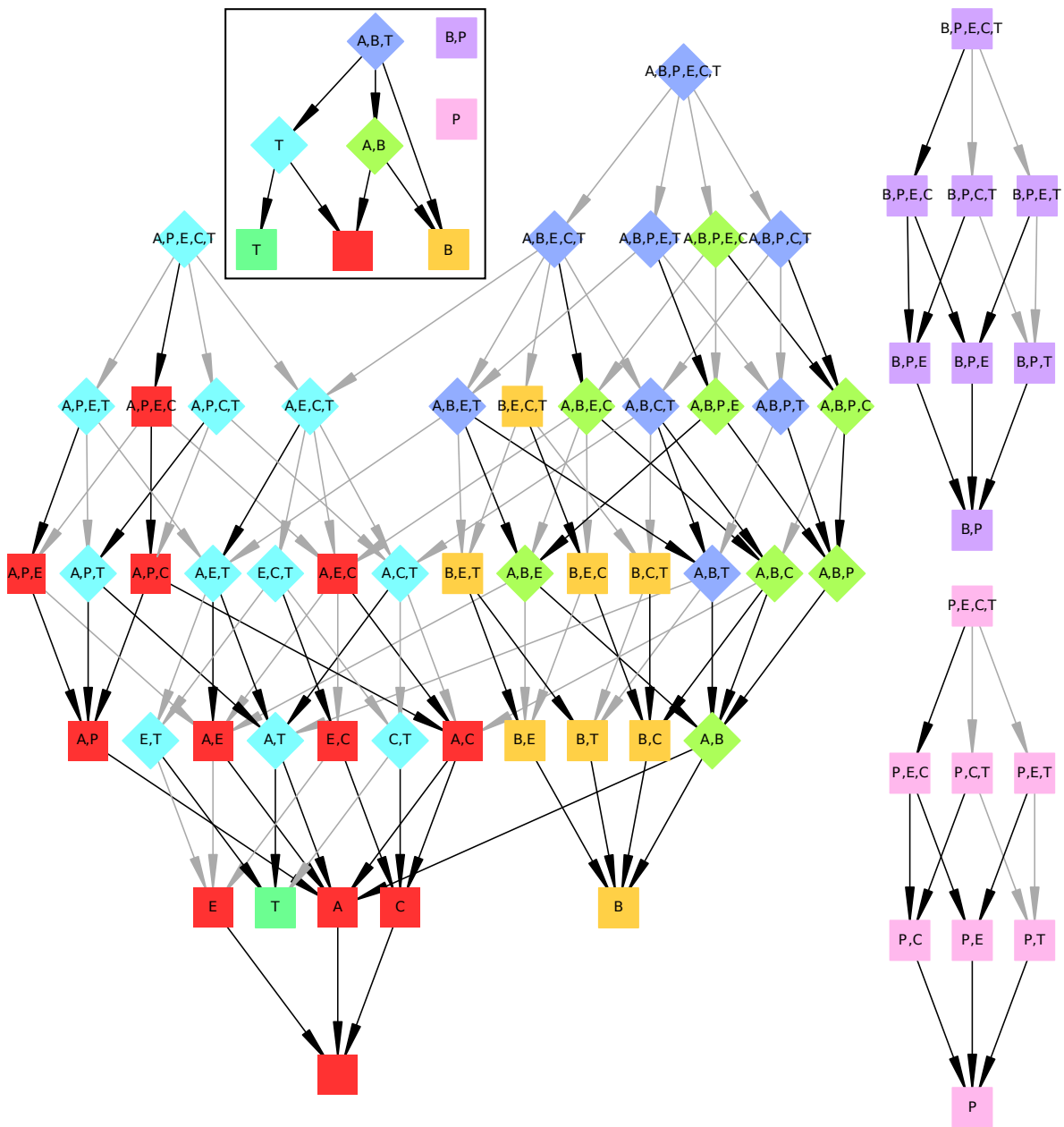


Figure 4: Complete graph of the 64 possible states for the system, with the transitions predicted by the model (defined in Fig. 3). Black (grey) arrows correspond to observed (unobserved) transitions in the experiments. States drawn with the same color have the same possible final states. Square states have only one possible final outcome; diamonds several ones. Inset: schematic (“merged”) graph of the complete state space, where all states with the same possible outcomes are merged into one super-node. Edges in this merged graph indicate that there exist transitions predicted between *some* of the states within super-nodes. Labels indicate which species are present in *all* the states of a given super-node.

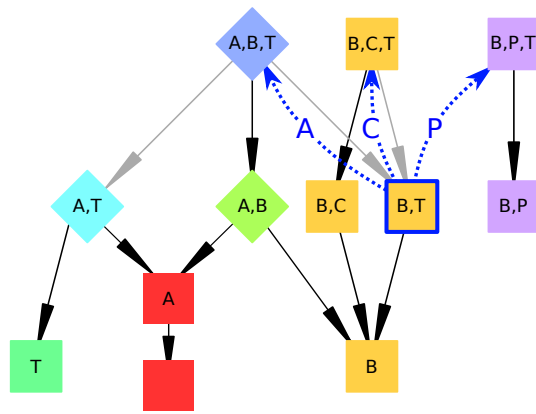


Figure 5: Subset of the state space (Fig. 4) corresponding to the invasion of state $\{B, T\}$ by species C, A or P (blue dotted arrows). Starting from $\{B, T\}$, invasion by C has a weak impact, invasion by P has a large and irreversible impact and invasion by A has an indeterminate effect (see text).