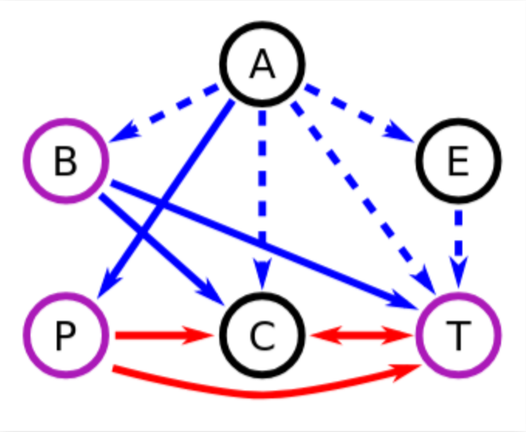
Inference of ecological networks and possibilistic dynamics based on Boolean networks from observations and prior knowledge

Loïc Paulevé, Cédric Gaucherel (AMAP, INRAe, CIRAD, Montpellier)

bioRxiv preprint: https://www.biorxiv.org/content/10.1101/2024.07.01.601264v1

Context, general motivation

- From observed (qualitative) change of species abundances, *generate hypotheses* on the underlying ecological network
- *Ecological network*: trophic (predation) + competition



- Network inference: take into account *prior expert knowledge*

Approach

Application:

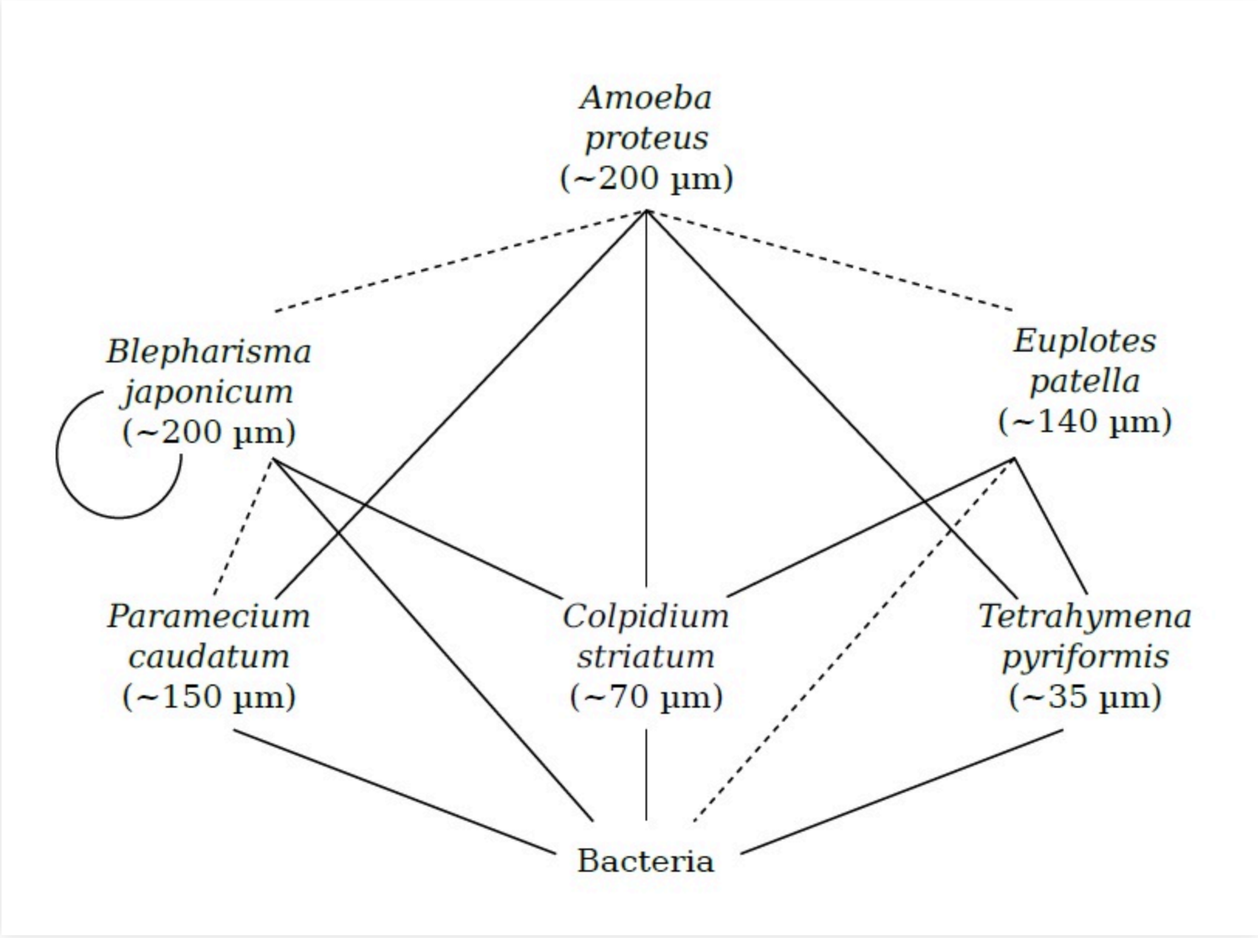
- Protist system with 6 species
- Extensive observations of *transitions* of species presence

Methodology:

- *Link ecological networks with influence graph* of Boolean networks
- Use *BoNesis* to *infer Boolean networks* being able to reproduce transitions
- Extract ecological networks from the influence graph
- Define *optimization criteria*: sparsity vs fidelity
- Apply with *different prior knowledge*, compare results

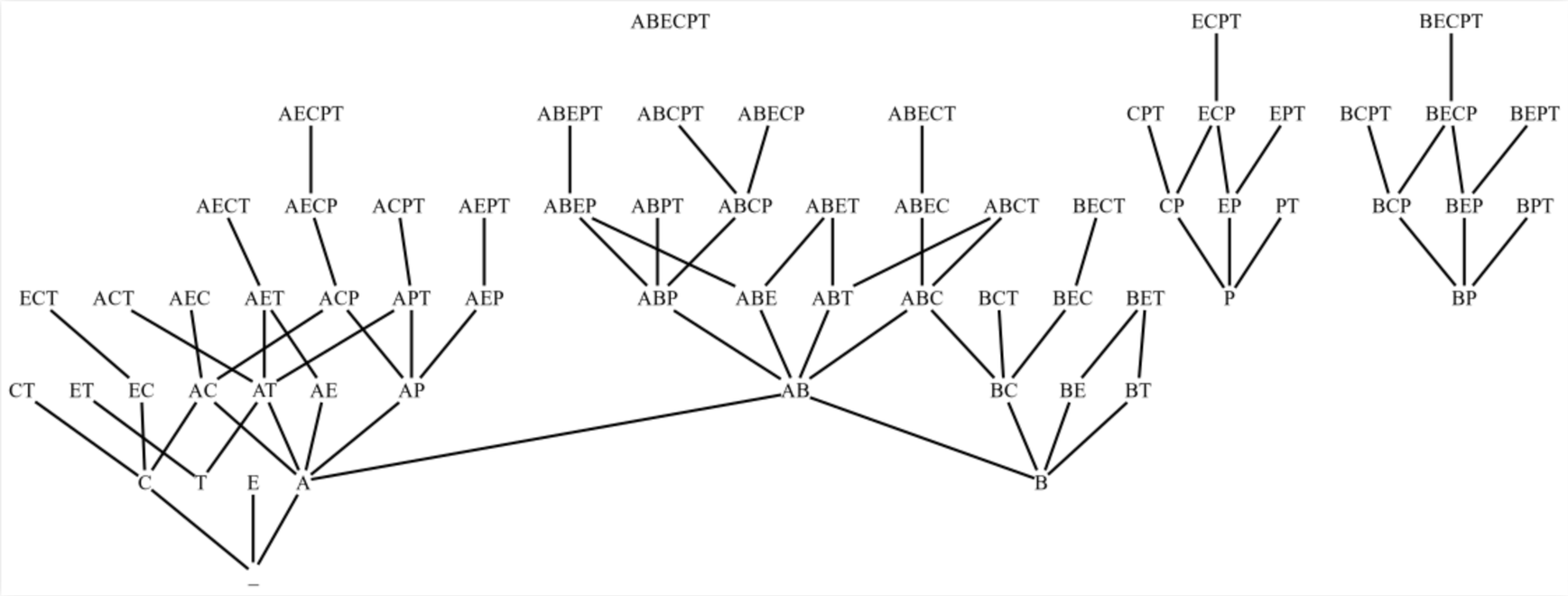
Protist community

Seminal protist experiment published by University of Sheffield (Weatherby, Law, Warran; 1998-2006):

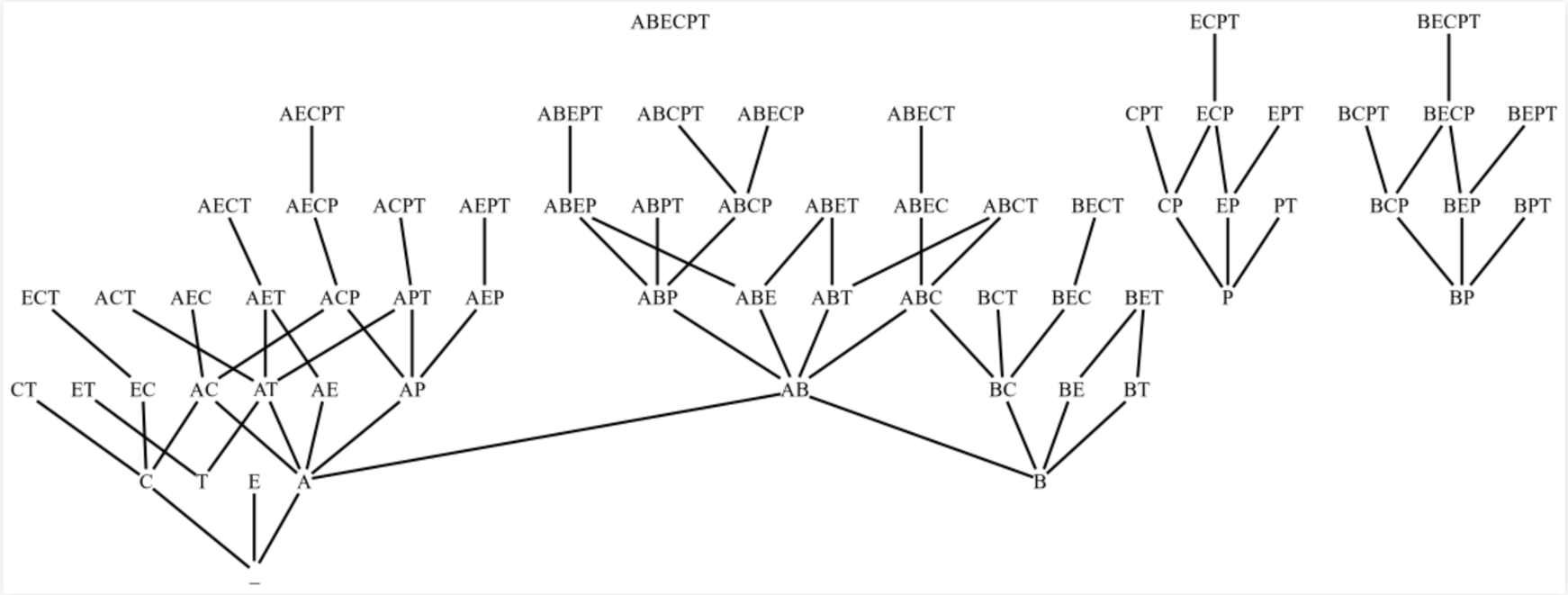


Observed transitions

- 63 combinations of protist species have been tested and replicated (6x)
- Bacteria always present
- Measurements in time series; qualitative analysis by M. de Goër de Herve (2022):



Dynamical properties

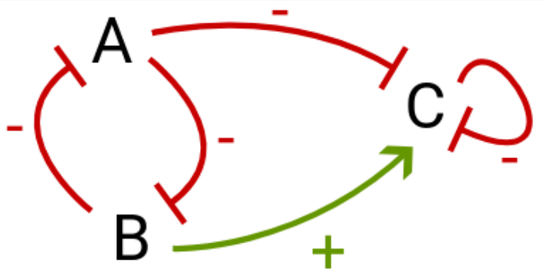


- Transitions are fully asynchronous (only one species disappears at a time)
- 4 steady states: _ (no survivor); B, P, BP
- From ABECPT (all), transitions are too fast to observe (uncertainty on possible transitions)

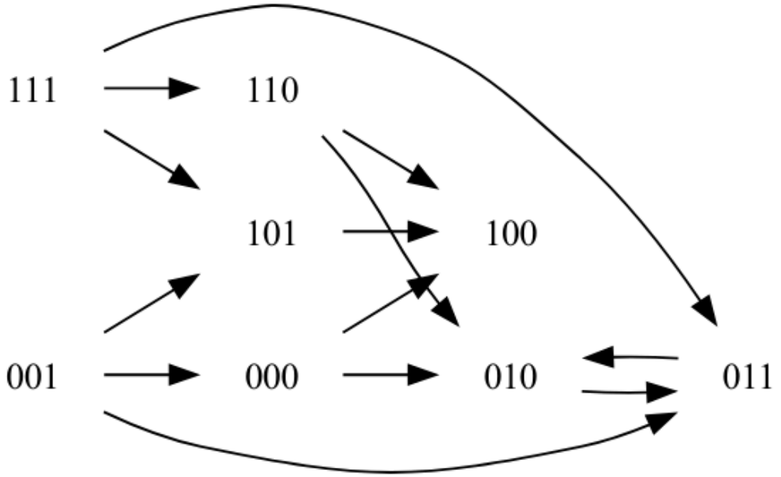
Boolean networks

$$f_1(\mathbf{x}) = \neg \mathbf{x}_2$$
$$f_2(\mathbf{x}) = \neg \mathbf{x}_1$$
$$f_3(\mathbf{x}) = \neg \mathbf{x}_1 \wedge \mathbf{x}_2 \wedge \neg x_3$$

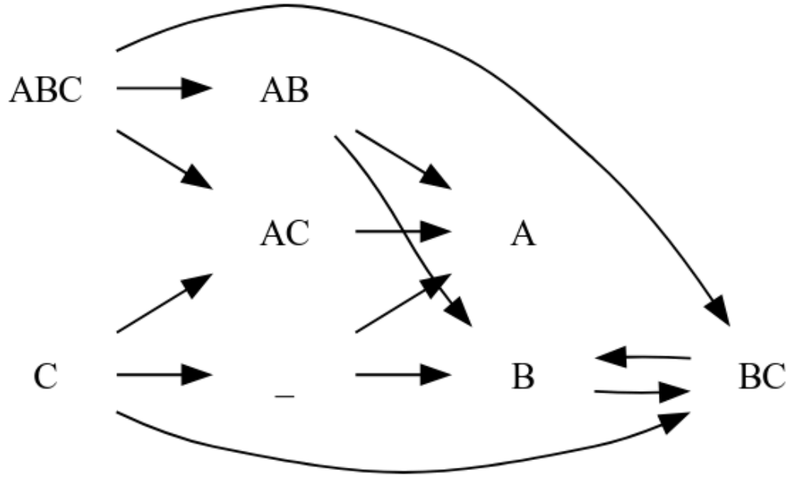
(a) Boolean network



(b) Influence graph



(c) Asynchronous dynamics with configurations as binary vectors



(d) Asynchronous dynamics with configurations as present species, assuming species A, B, C have index 1, 2, and 3, respectively

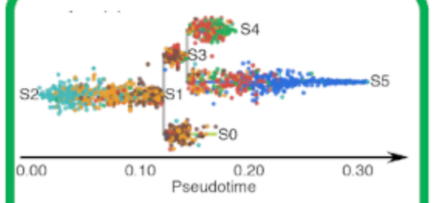
General objective for the inference

What do we want:

1. Boolean networks that are able to reproduce *at least* the observed transitions
   ↪️ find logic of presence/absence for each species that can explain the transitions
2. *minimize deviations* from observed transitions
   ↪️ models that predict as less as *unobserved transitions* as possible
3. sparse models: smallest (nb of edges) ecological networks

BoNesis

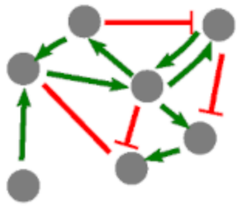
Experimental data



scRNA-seq
bulk time series
...

+

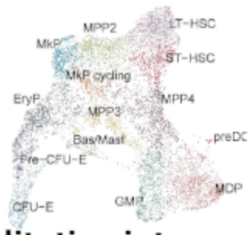
Knowledge



TF-TF interaction DBs
Pathways
...

+

Biological expertise



Qualitative interpretation
Markers
...

BoNesis - generic BN synthesis engine

- Symbolic Artificial Intelligence, automatic reasoning (ASP)
- Scalable to real-world data (X,000 genes/cells)
- Open source (CeCiLL/GPL)

Ensembles of Boolean networks

Predictions

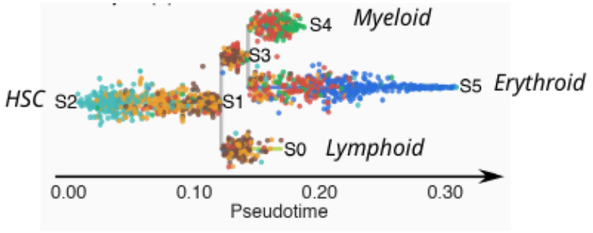
Gene selection
Key regulations
Mutations

BoNesis • Proof of concept on scRNA-seq data

Modeling hematopoiesis cellular differentiation processs

PhD work of
Stéphanie Chevalier

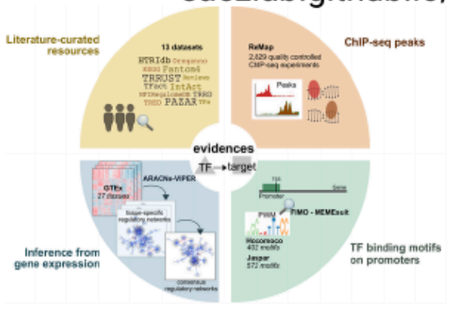
scRNA-seq data



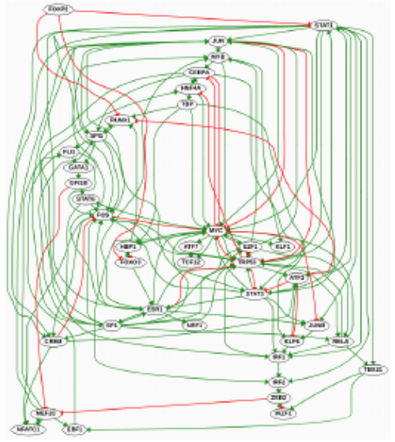
~4,800 nodes (proteins)
~16,000 influences

DoRothEA TF-TF db

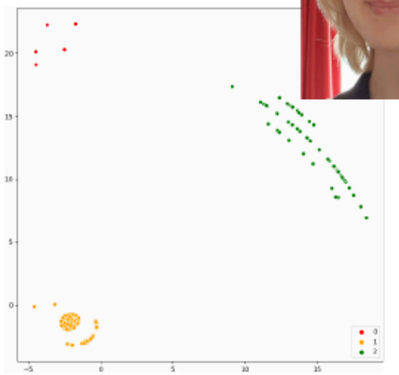
saezlab.github.io/dorothea



1.5 million interactions between
1399 TFs and 27,976 target genes

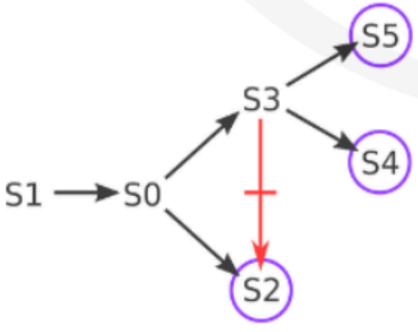


Node selection



Models clustering

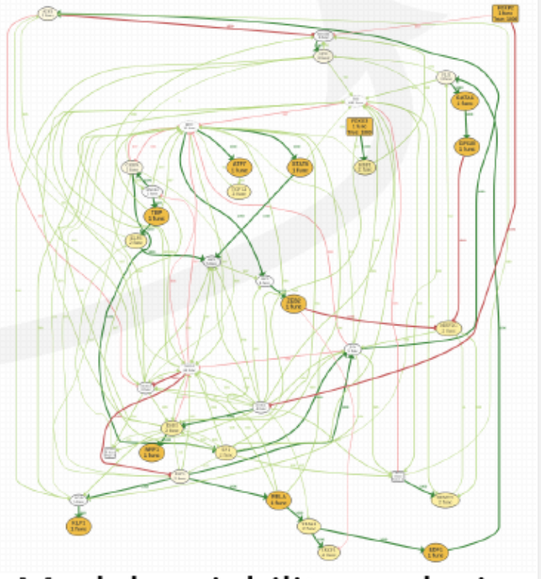
Boolean dynamical properties



BoNesis

$$f_1(x) = \dots$$
$$\dots$$
$$f_n(x) = \dots$$

Solution sampling



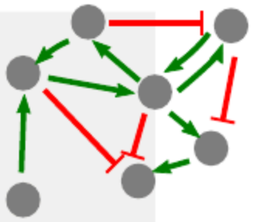
Model variability analysis

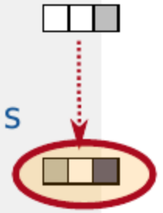
BoNesis in a nutshell

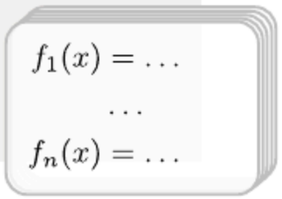
https://**bnediction.github.io/bonesis**

```
dom = bonesis.InfluenceGraph(...)
data = {
 "init": {"A": 0, "B": 0, "D": 0},
 "marker": {"A": 1, "C": 1}
}
bo = BoNesis(dom, data)
x = ~bo.obs("init") # cfg matching with init
y = ~bo.obs("marker") # cfg matching with marker
bo.fixed(y)  # y is a fixed point
x >= y   # exists trajectory from x to y
x >> "fixedpoints"  ^ {bo.obs("marker")}
   # each fixed point reachable from x matches
   # with marker (have A=1 and C=1)

# enumerate solutions:
for f in bo.boolean_networks():
     ....
```







Modeling

❶ Domain of Boolean networks

➜ influence graph / partially specified BN

➜ single BN (model checking/control)

❷ Observations binarized

➜ can be partial

❸ Dynamical properties

➜ link configurations with observations

➜ trajectories, fixed points, trap spaces, mutants, ...

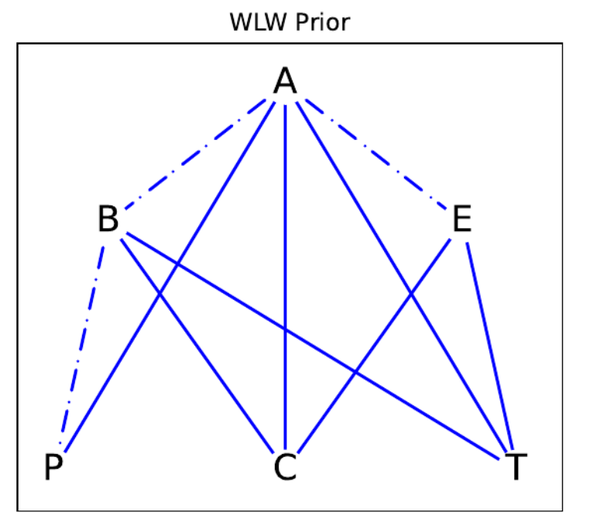
Solving

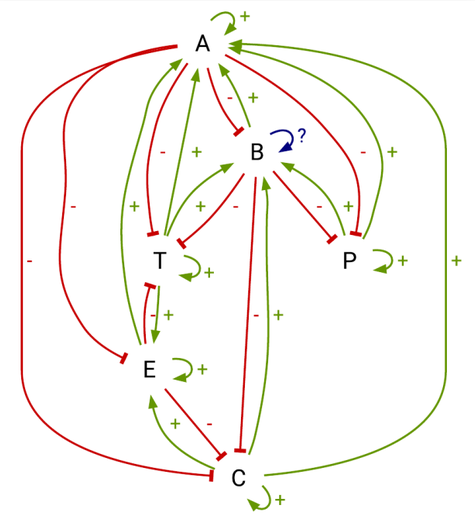
- Optimizations (BN size, error, ...)
- Different views of solutions

➜ Boolean networks (w/ or w/o diversity),

➜ Influence graphs, configurations, ...

Ecological networks vs Boolean networks





- A predates B means A has a negative influence on B, and B a positive on A
- positive loops: mostly to allow for persistent disappearance (e.g., $f_A(A=0, C=0) = 0$ and $f_A(A=1, C=0) = 1$)
- non-sufficiency of prey (dashed edges): deny some logic functions (e.g., $f_A(x) = A \wedge E \wedge \square$)

Encoding in BoNesis

1. Domain of candidate Boolean networks:
    - either complete graph (zero knowledge), or subgraph of prior knowledge
    - additional structural constraints (acyclicity, sufficient trophies, …)
2. Transitions that must be reproduced (fully-asynchronous)

```
with bo.scope_reachability(max_changes=1):
    for x, y in obs.edges():
        ~bo.obs(cfg_of_present(x)) >= ~bo.obs(cfg_of_present(y))
```

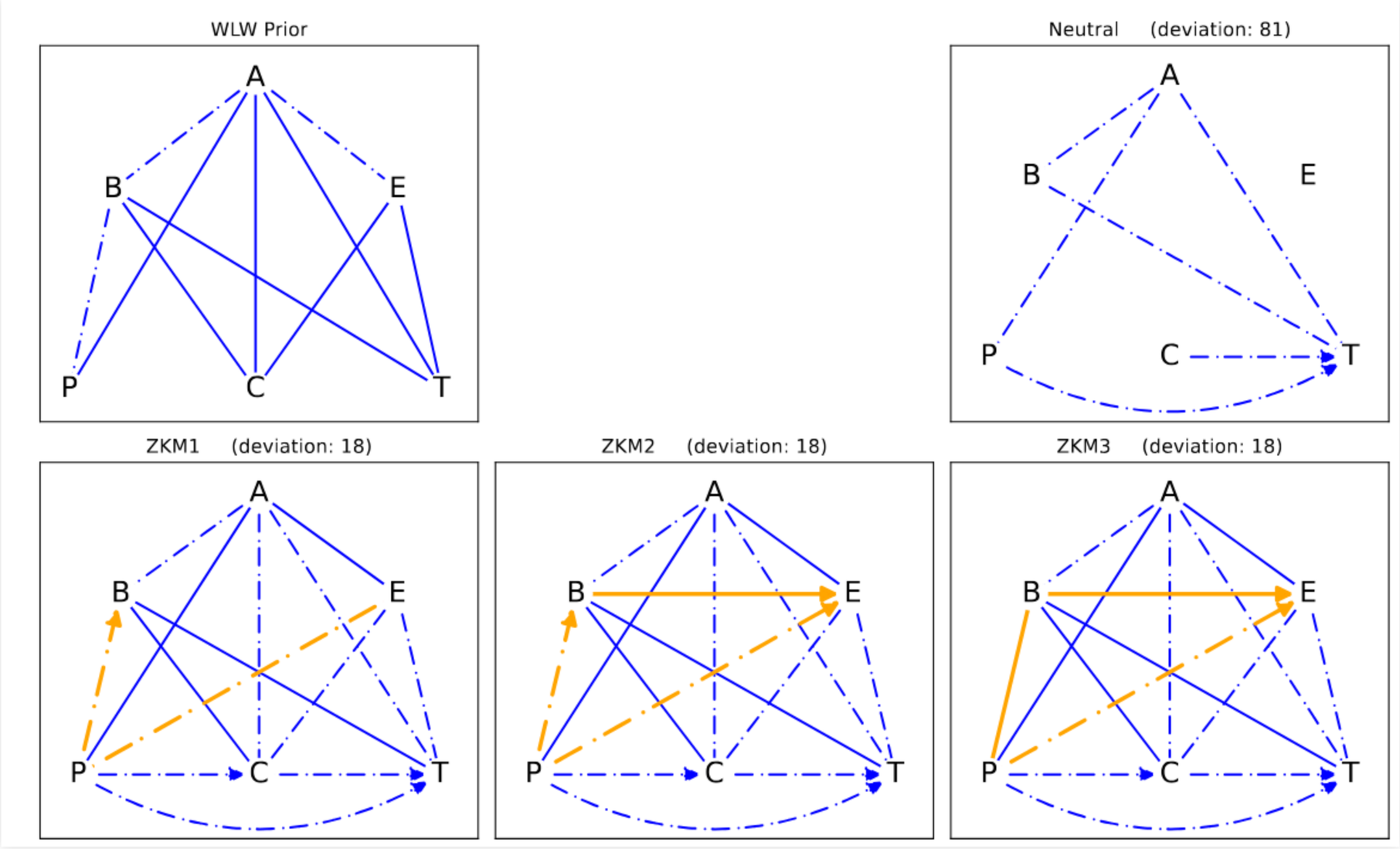
3. Optimization criteria (minimize non-observed transition + size of ecological network)

Protists case study, part 1: inference from zero knowledge

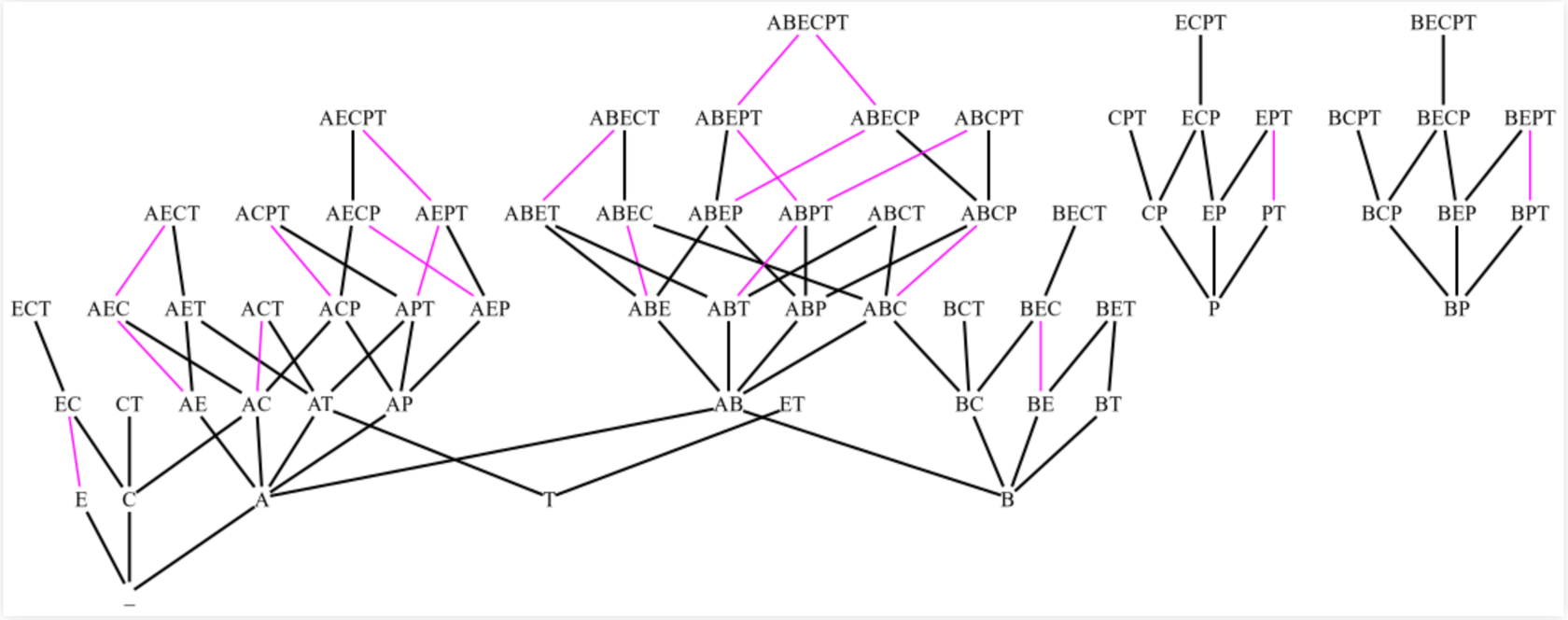
Let’s assume nothing on the ecological network… except it must be acyclic

Zero-knowledge inference

Neutral: minimal model reproducing the observations; ZKMx: min deviation



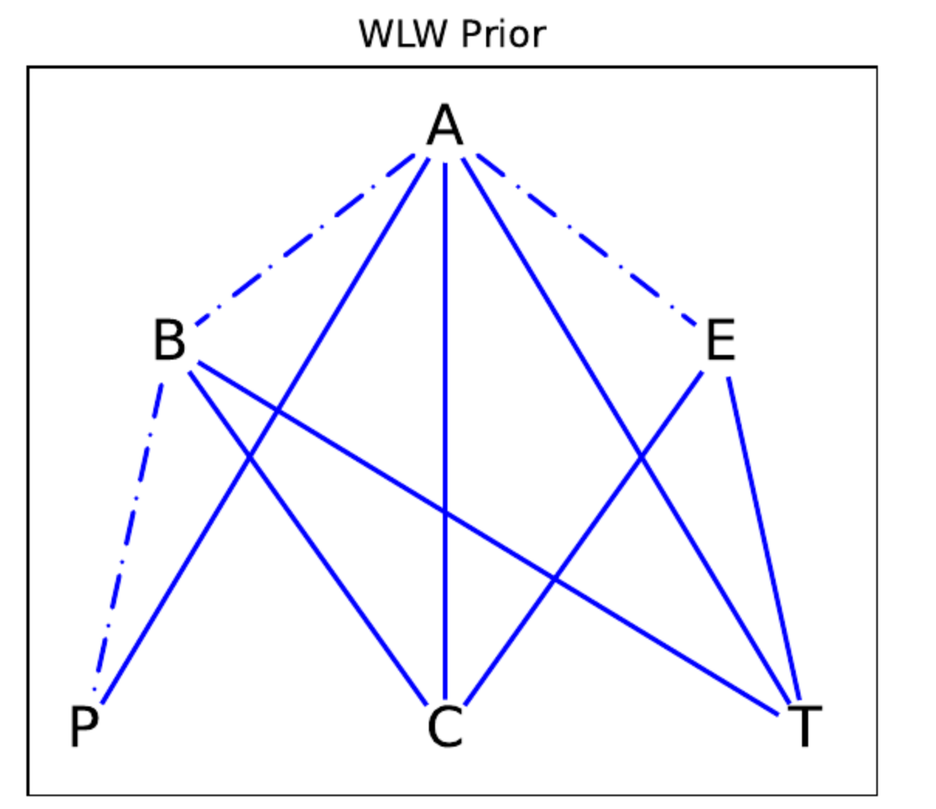
Transitions of ZKM3



Protists case study, part 2: integrating prior knowledge

- 💡 only employ ecological interactions deemed possible by an expert

Prior trophic network from Weatherby, Law, and Warran

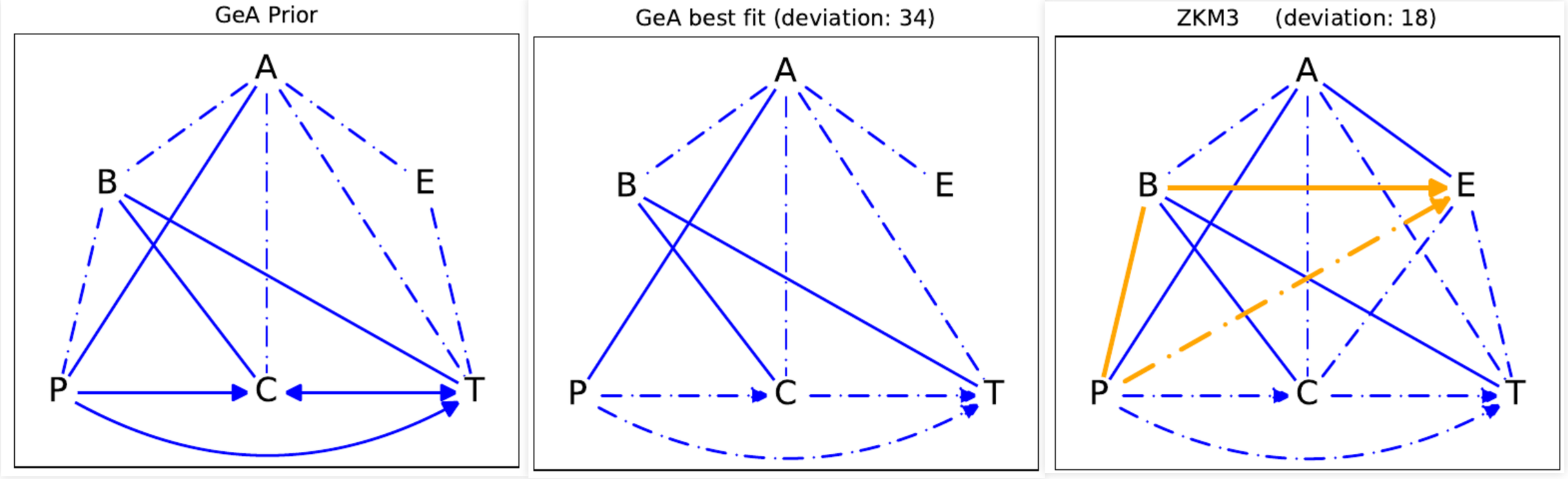


❌ UNSAT: T can survive alone, but not in presence of C or P

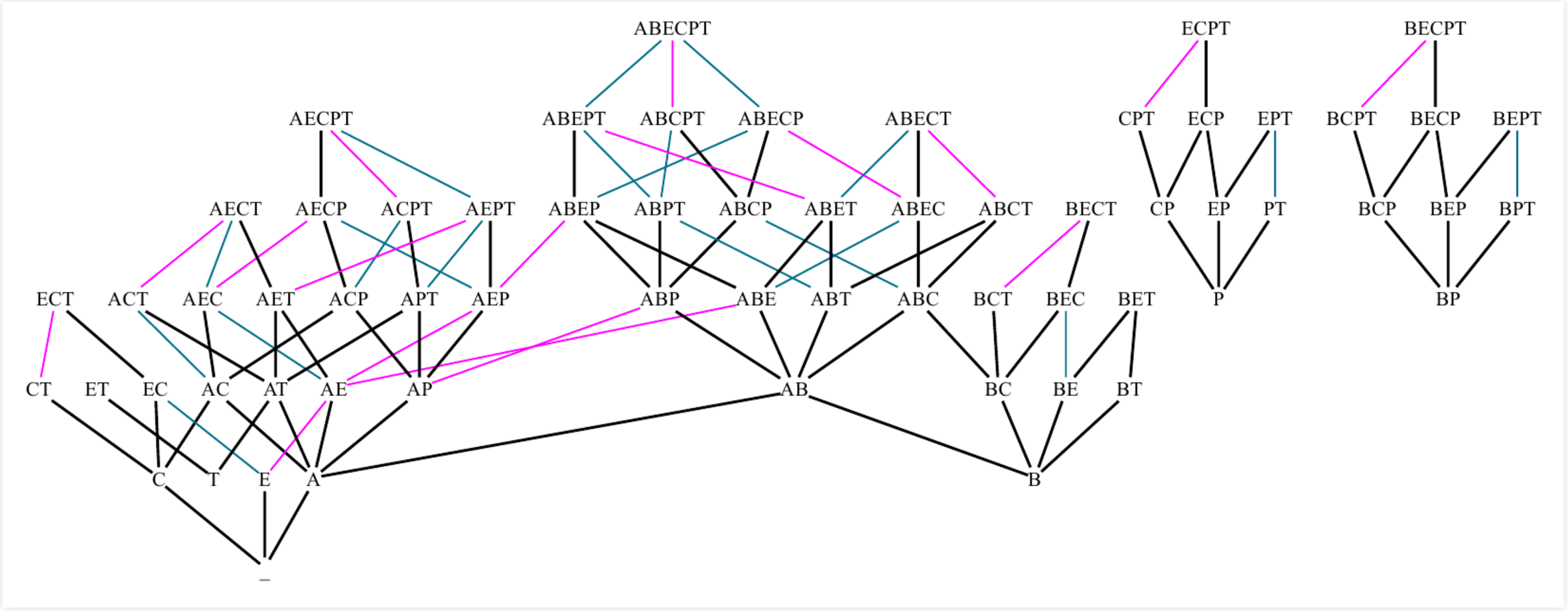
↪️ Network is incomplete

Extended prior trophic network

De Goër de Herve et al. (2022) proposed an extended trophic networks, coupled with potential competitions



… resulting transition graph



Discussion

Inferred ecological networks:

- Zero-knowledge models are already quite plausible (ZKM3 fully reflect the size of the species)
- One can prove that prior knowledge can be incomplete (and point where)

Inferred dynamical models

- Added transitions often complete *diamonds* (concurrent transitions)
- Less deviation (spurious transitions) = more complex logic/network
  ↪ trade-off with the abstraction level: time/speed could also explain absence of transitions
  ↪ importance of prior knowledge to *prevent overfitting*

Discussion

Pros:

- Easy to implement and quick to execute
- Flexibility on the specification/constraints

Limits / applicability to other case studies

- Small system; with large systems, we need to bound the complexity of the Boolean functions
- Competition vs predation can be tricky to distinguish from Boolean rules
- Would be nice to try other systems / compare with other tools (LFIT?)
- BoNesis supports reachability, steady state, and control properties: handle more usual ecological data?
- BoNesis restricted to locally-monotone BNs; WiP to non-monotone