

1.  
methods for the  
**inference** of  
**co-occurrence**  
**networks**

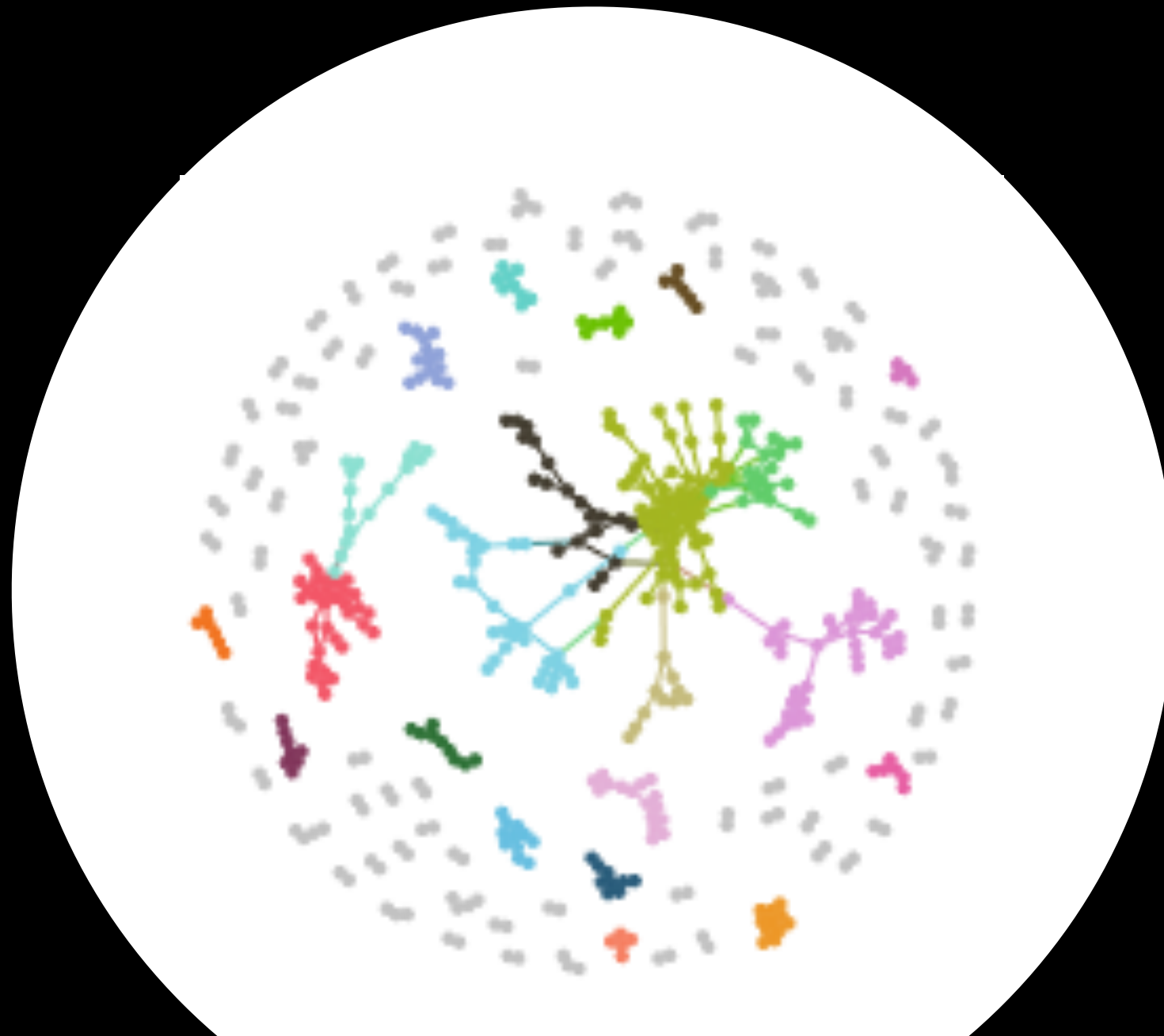
(a critical assessment)

2.  
**co-occurrence**  
**links** from  
**samples:**  
close to  
**functional?**

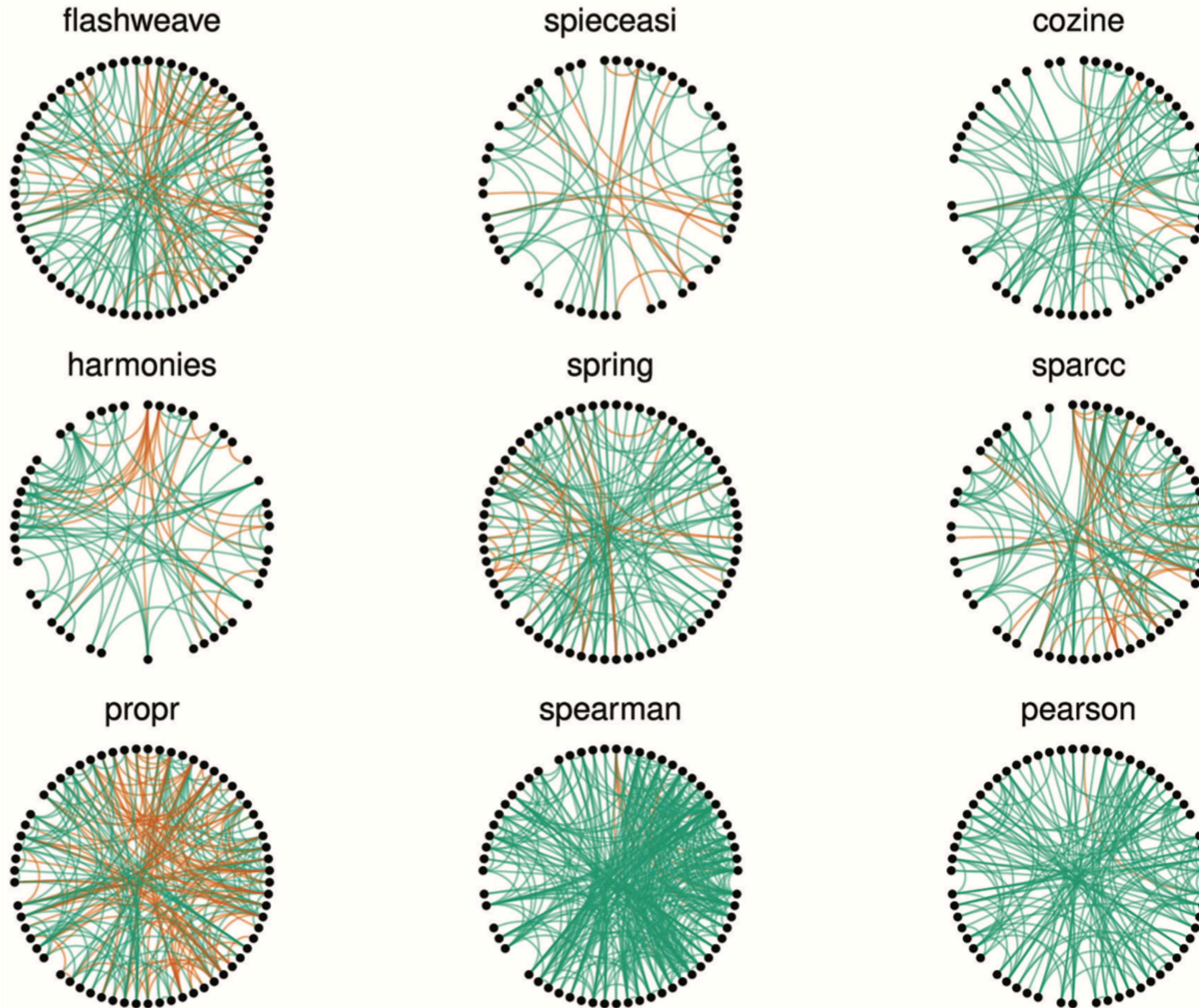
(early results on biota)

3.  
(future outlook:)  
**inferring**  
**functional**  
**networks**  
**we can act**  
**on**

# 1. methods for the **inference** of **co-occurrence** networks (a critical assessment)



# Many methods, little convergence



Inferring microbial co-occurrence networks from amplicon data: a systematic evaluation. J. MSystems, 2023.

(Nematodes, sand, filtered,  
CCS-normalised abundance)



Pearson/etc. corr.,  
SparCC/CCLasso/etc.

assume  
**independent,**  
**pairwise,**  
**undirected**  
relationships

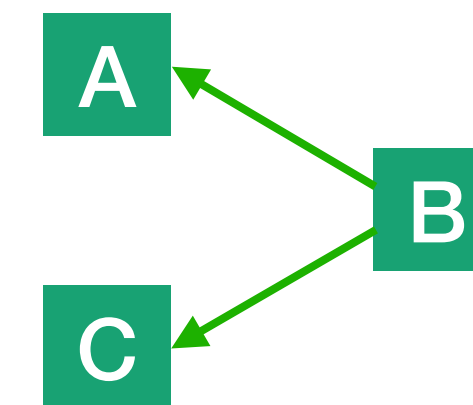


(and none is  
entirely true)

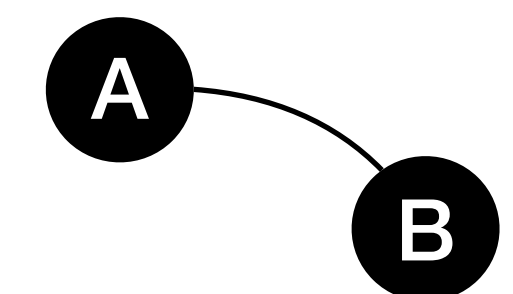
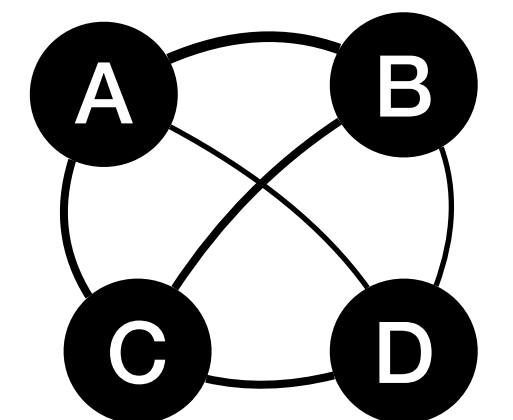
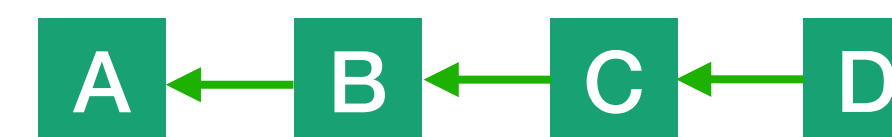
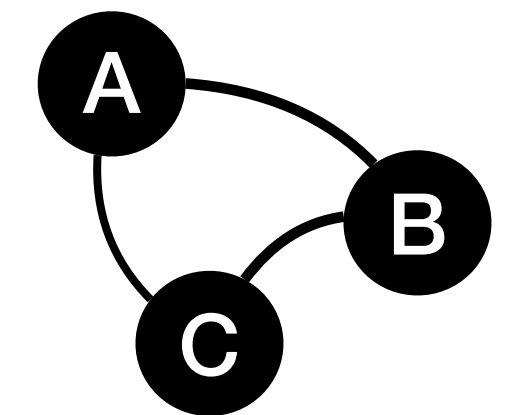
On the **pairwise** assumption (more  
fundamentally): see  
Challenge #5: Higher-order interactions (HOIs),  
**Open challenges for microbial network  
construction and analysis, ISME 2019.**

**dense,**  
**spurious,**  
**unstable**  
networks

ground truth

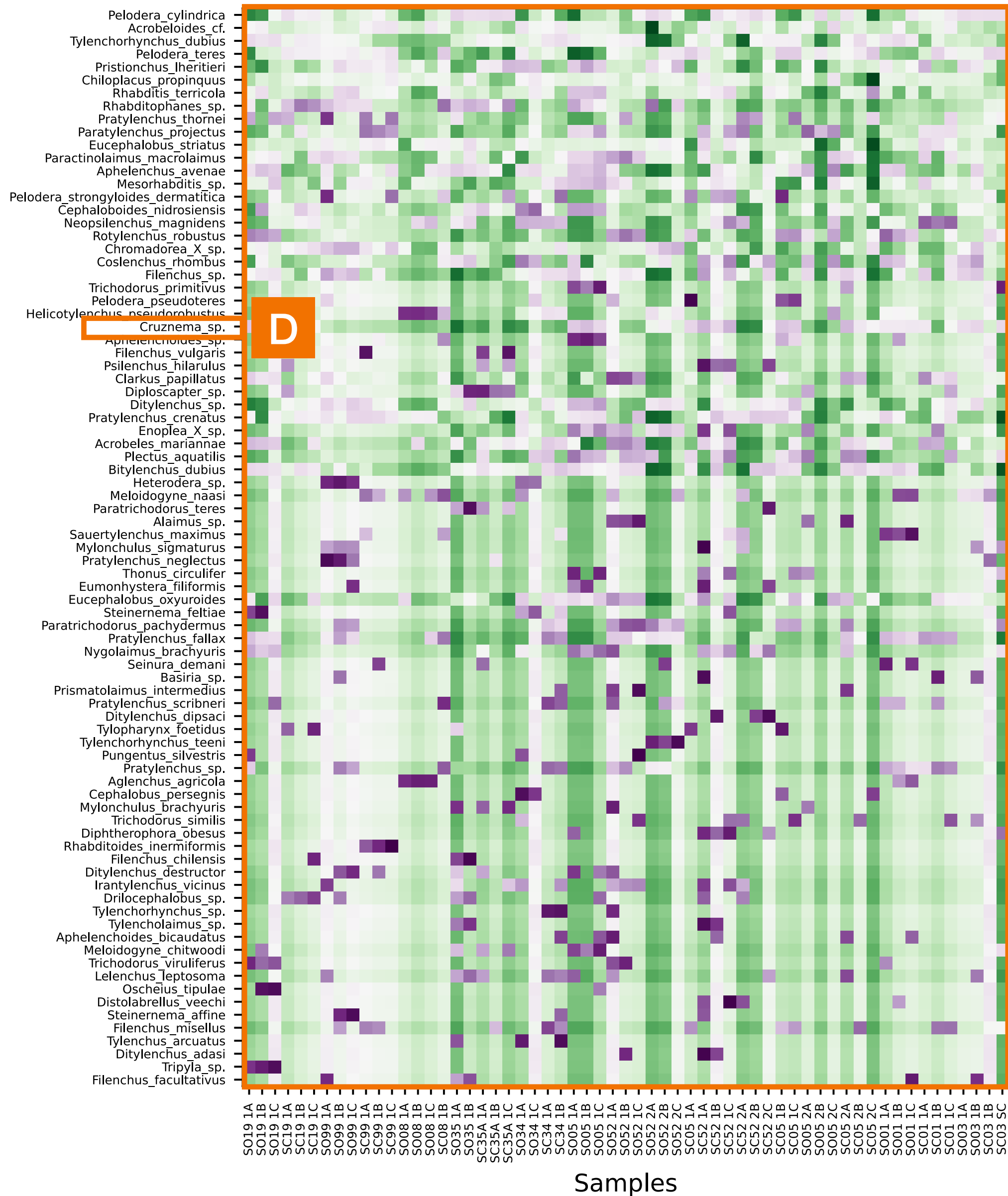


co-occurrence



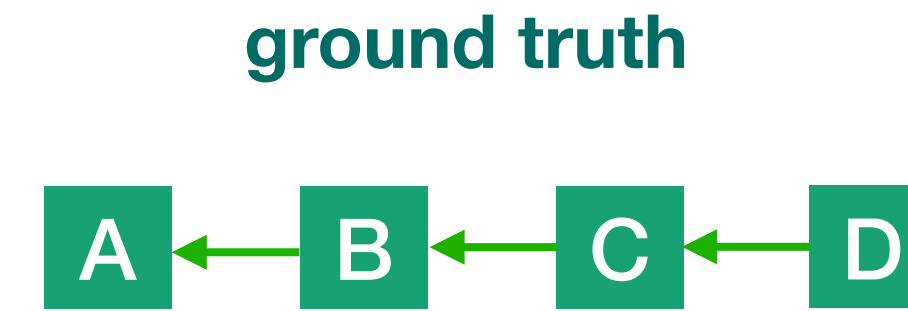
**thresholding?**

(Nematodes, sand, filtered,  
CLR-normalised abundance)

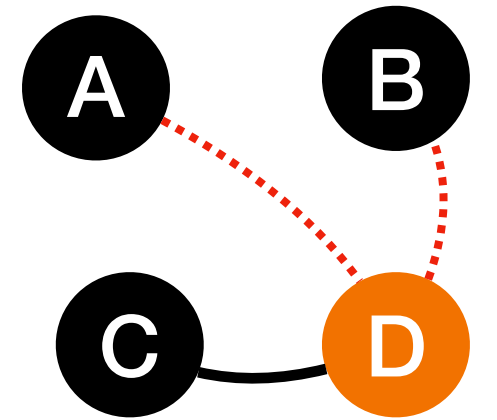


Spiec-Easi, EleMi  
lift the assumption of  
**independent**  
relationships

(they differ in robustness and  
sparsity, because they use  
different regressors, with  
weaknesses)



co-occurrence  
neighbourhood of  
**D** only (repeat for  
all taxa)



taxa whose abundances  
**do not help regress D's**  
abundance any better  
than C

...have their links to D cut!

(a **global optimisation**  
problem)

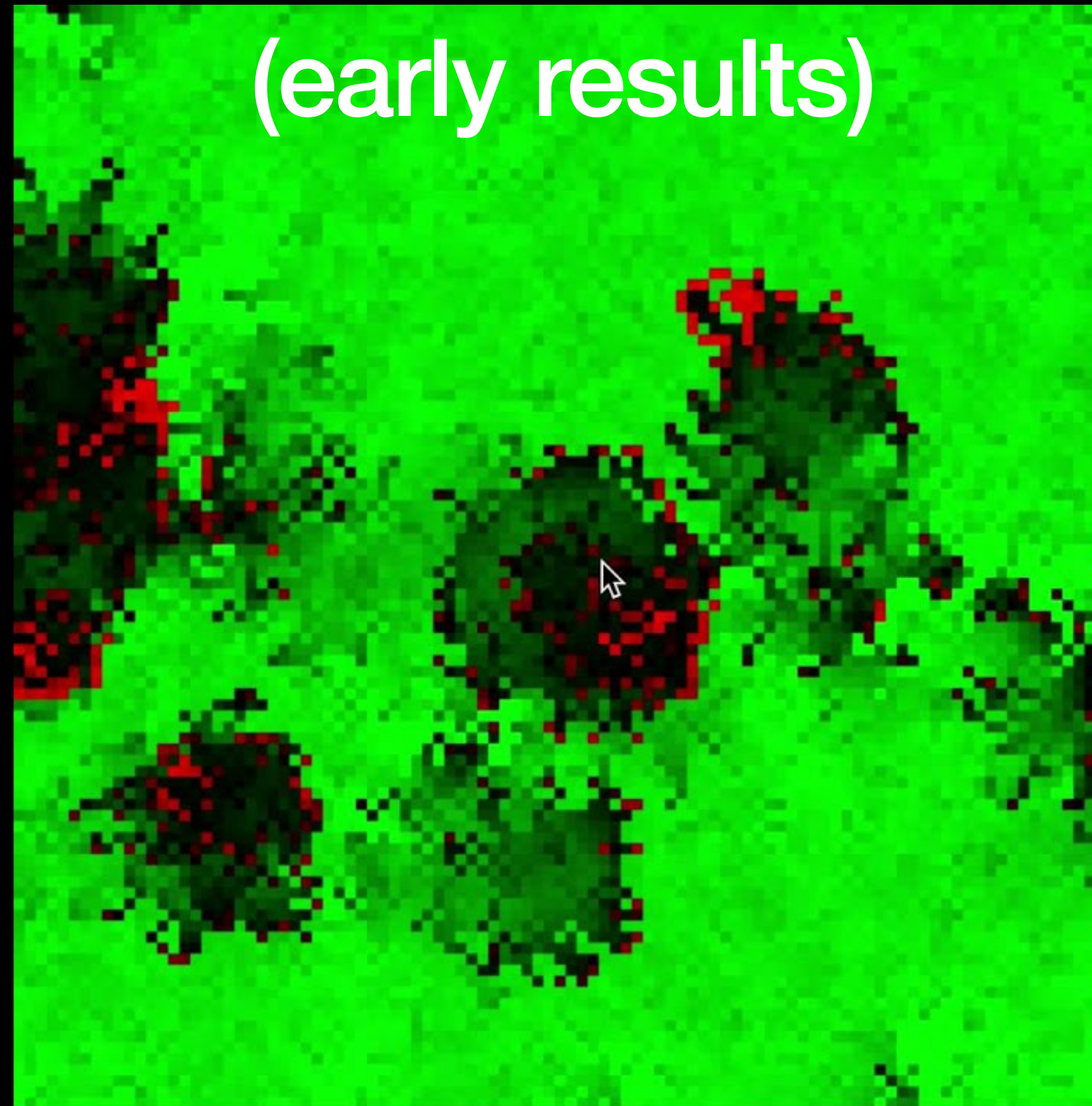
EleMi: A Robust Method to Infer Soil  
Ecological Networks with Better  
Community Structure.

Chen, Bucur. Int. Conf. Complex  
Networks (2024)

# 2.

## co-occurrence links from samples: close to functional?

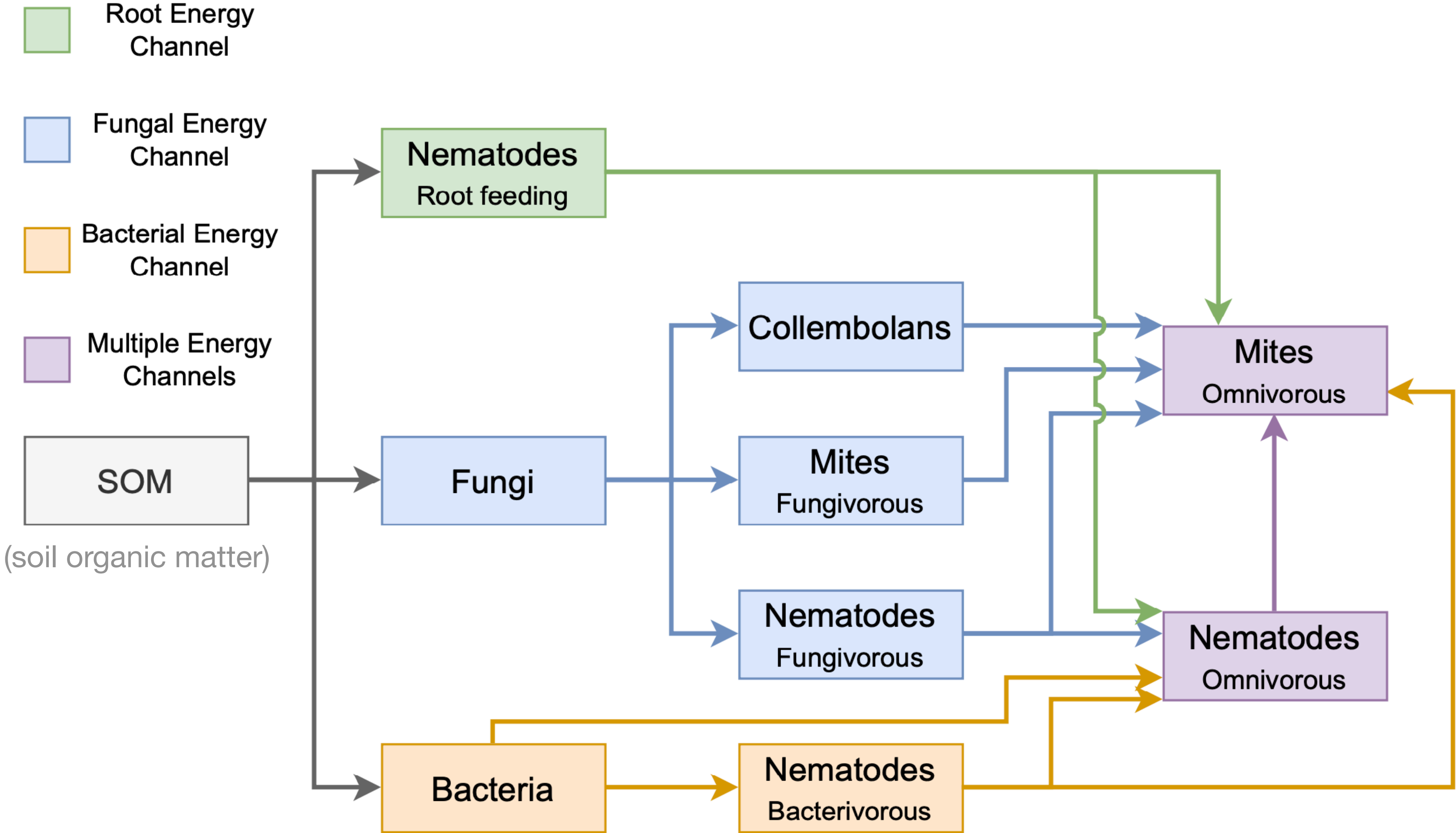
(early results)



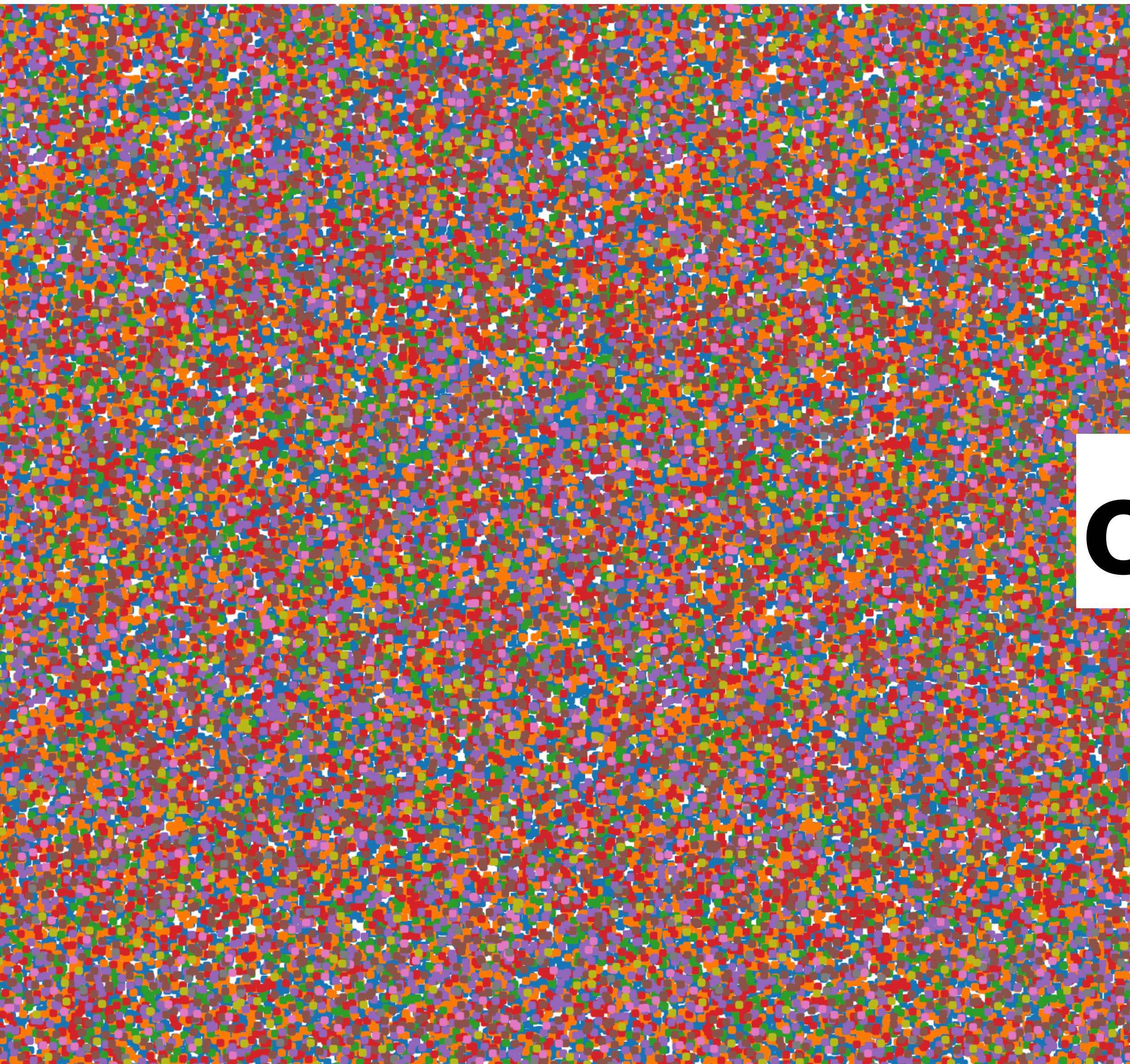
Community in a Petri dish,  
simulation:

[https://www.youtube.com/  
watch?v=LwLFw1\\_GGnU](https://www.youtube.com/watch?v=LwLFw1_GGnU)

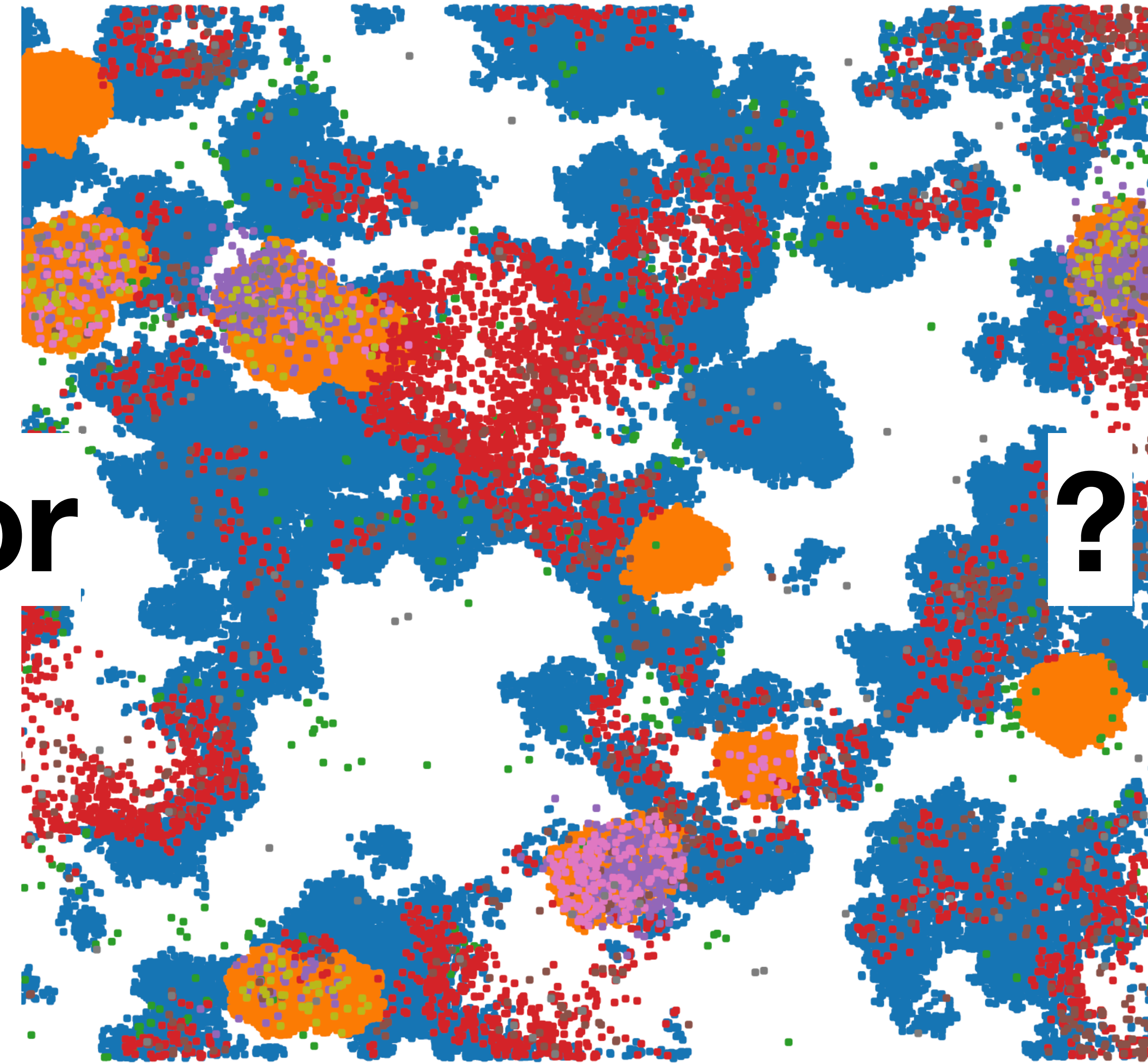
As a case study for simulation, take this **trophic model**, with big-picture “taxa”:



2m x 2m soil plot seen from above, dot=taxon (like a Petri dish)



or

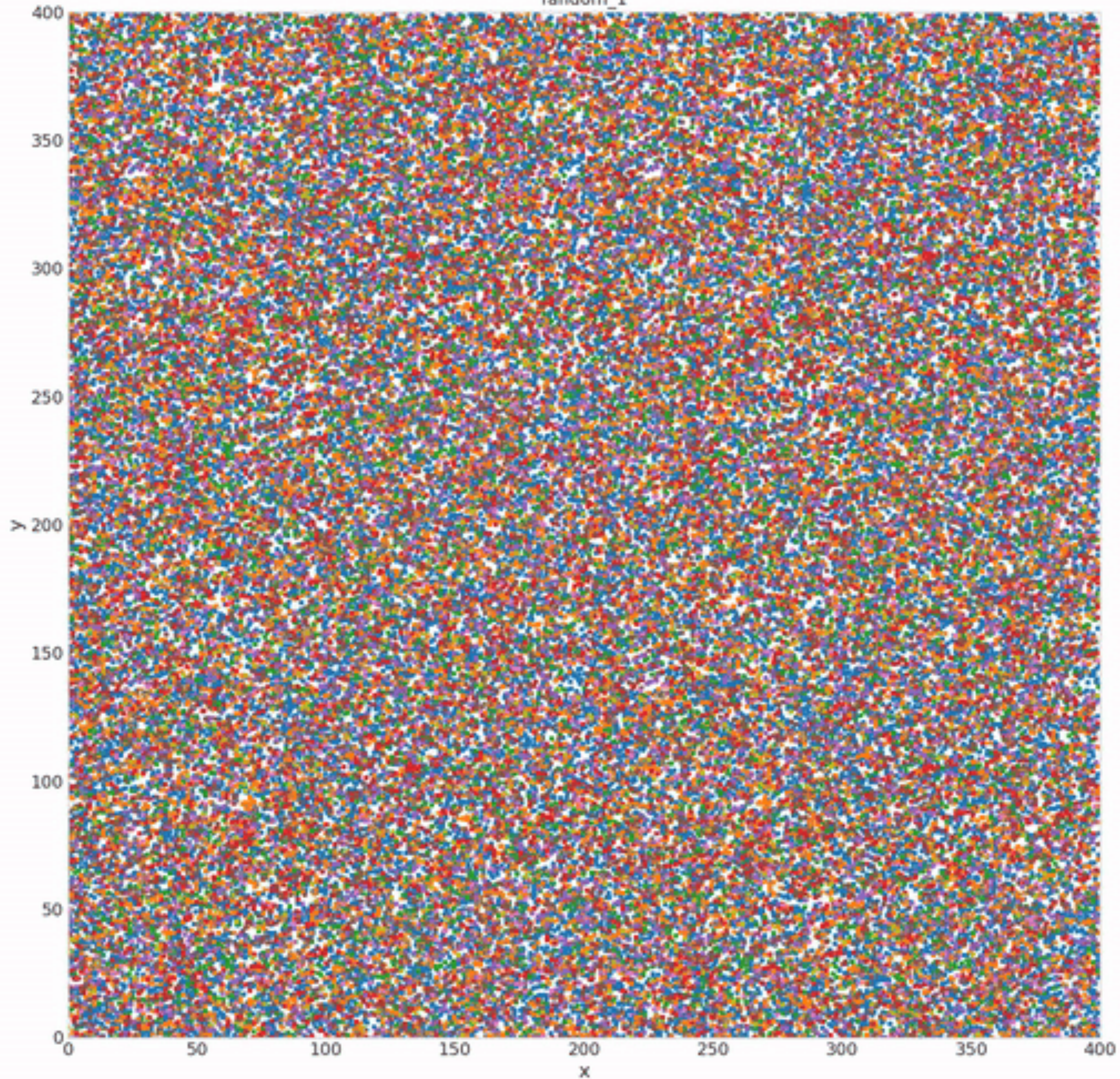


?

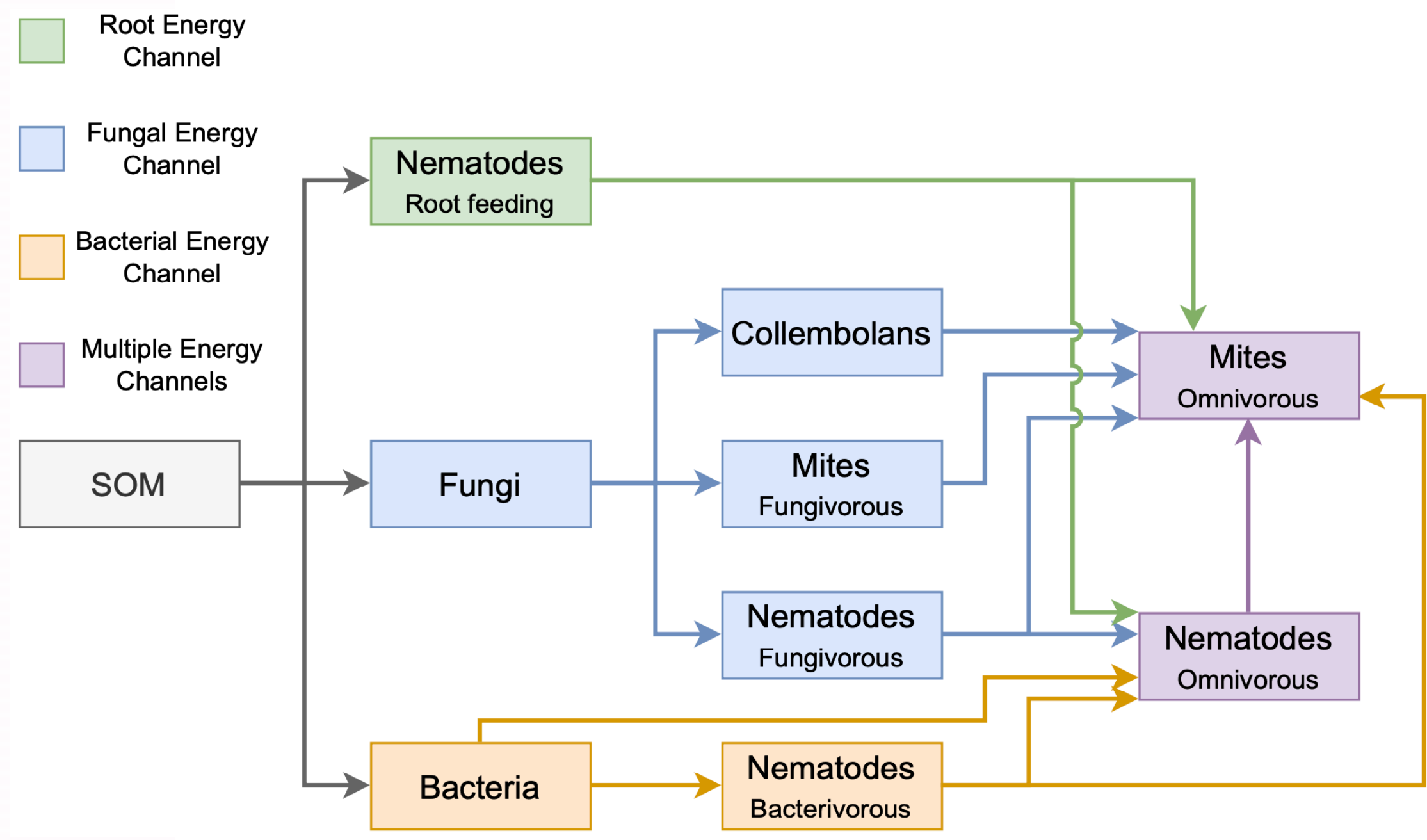


random\_1

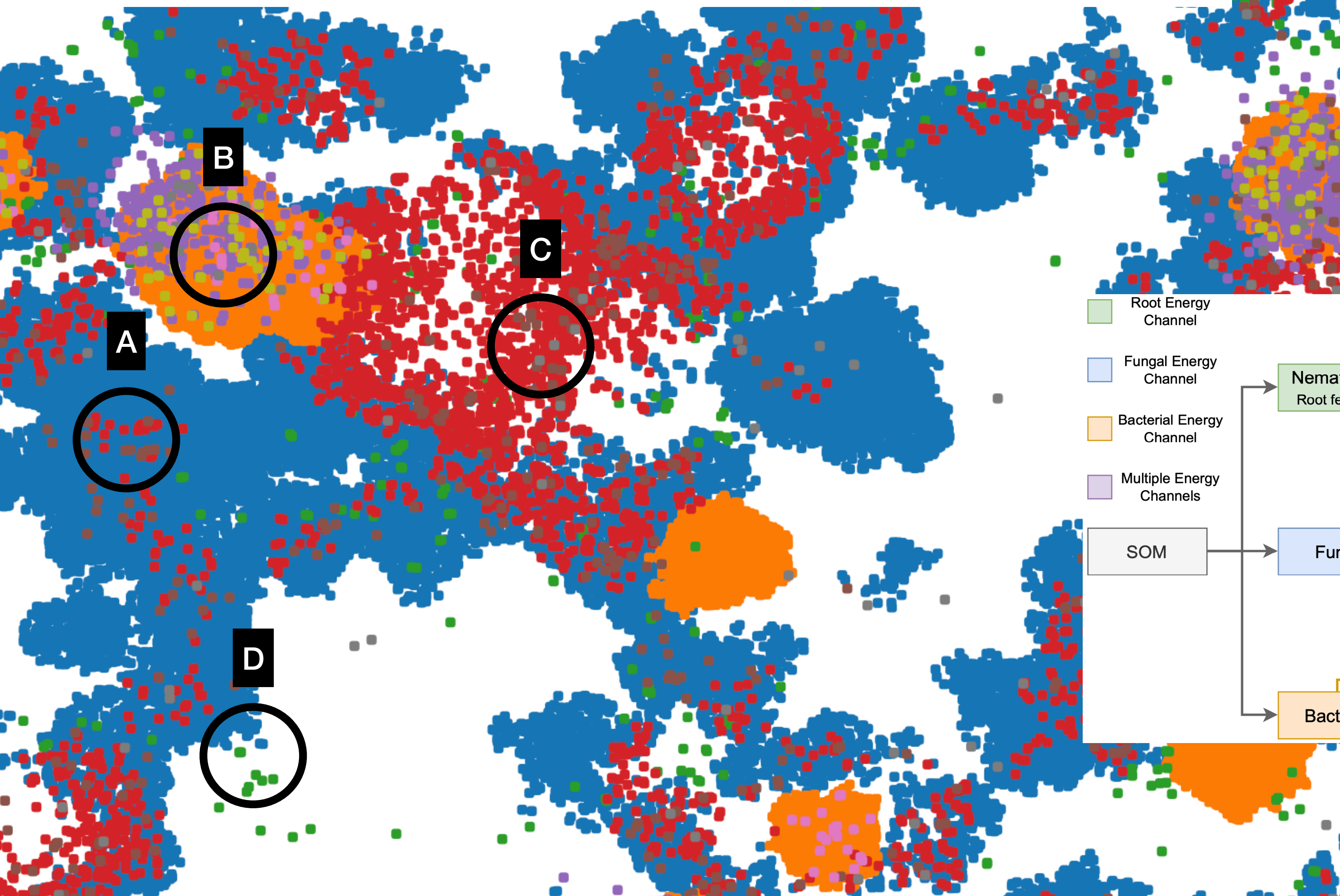
Movie at <https://github.com/timovdk/BLOSSOM>



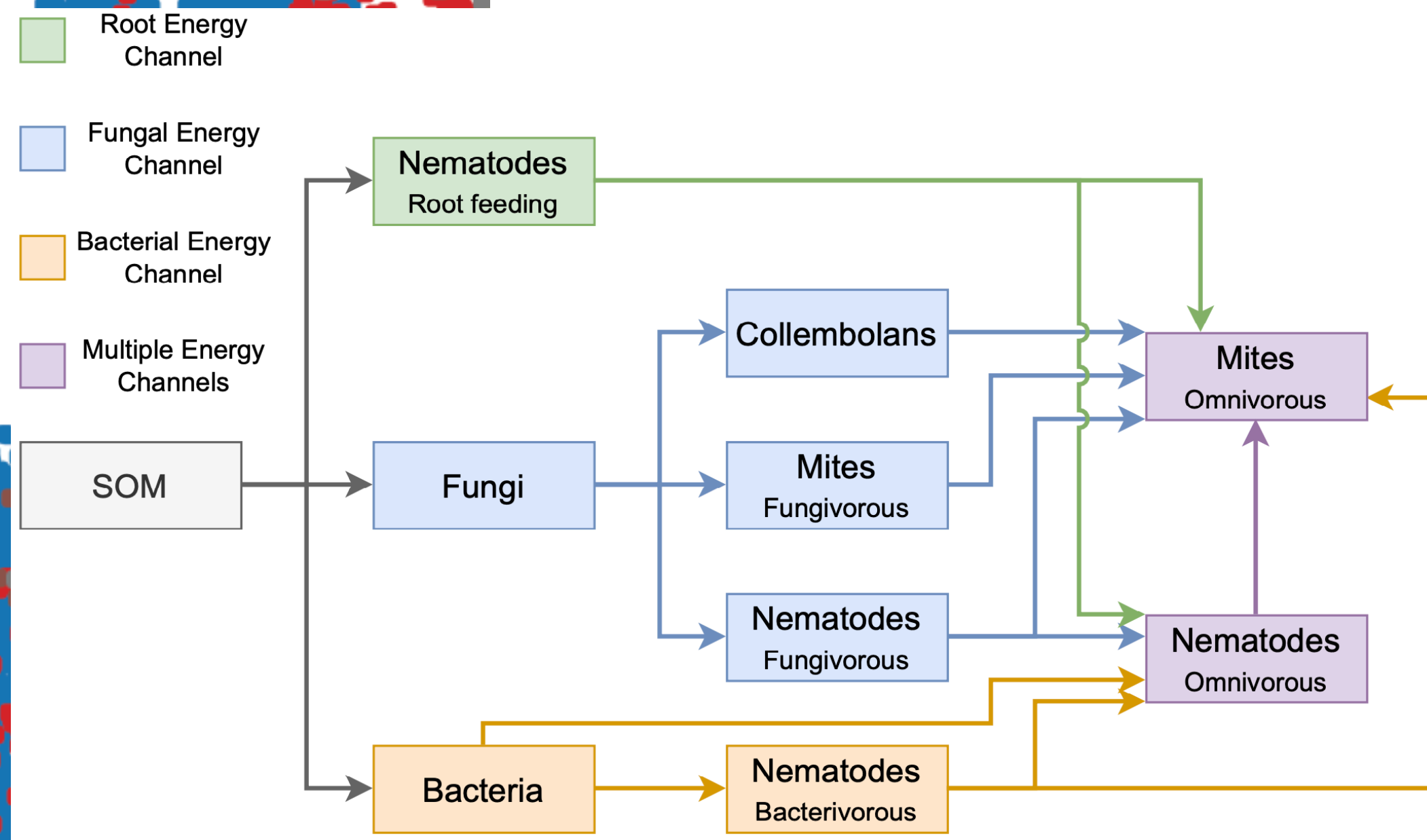
- Bacteria
- Fungi
- Root-feeding Nematodes
- Bacterivorous Nematodes
- Fungivorous Nematodes
- Omnivorous Nematodes
- Fungivorous Mites
- Omnivorous Mites
- Collembolans



# Can samples reconstruct **all** functional links? With **direction**?



- Bacteria
- Fungi
- Root-feeding Nematodes
- Bacterivorous Nematodes
- Fungivorous Nematodes
- Omnivorous Nematodes
- Fungivorous Mites
- Omnivorous Mites
- Collembolans



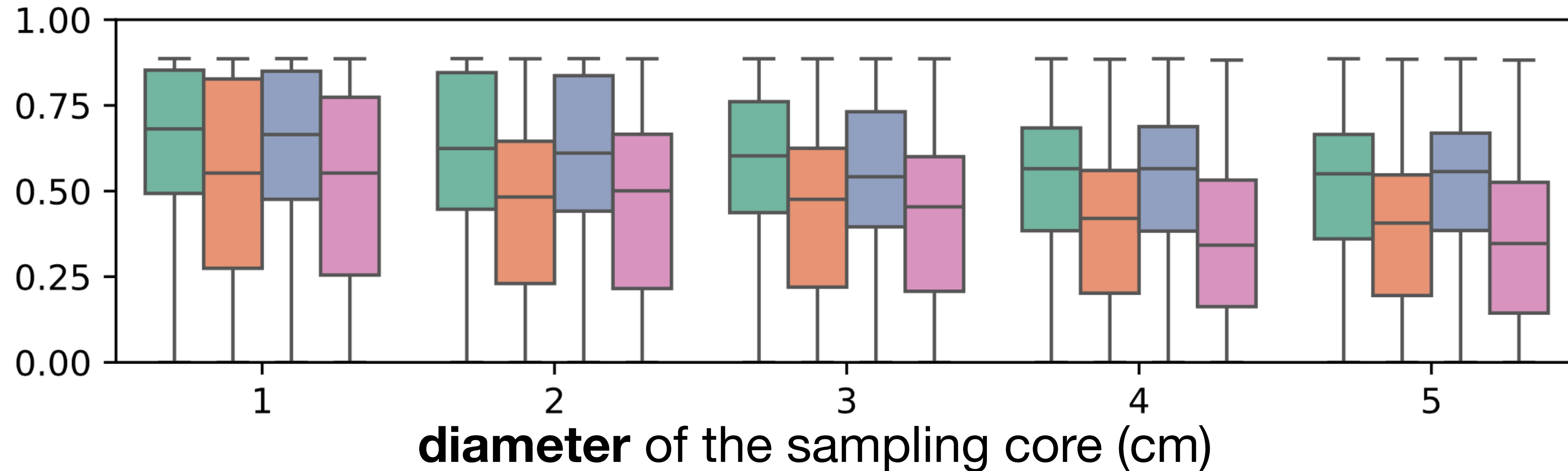


## When pooling 12-16 cores from 2 m x 2 m plots:

mean absolute  
**error** of  
co-occurrence  
**weights**  
are high!

(between  
**samples** and  
the whole **plot**)

**big outliers** for  
the rare taxa!



See also: **How sample heterogeneity can obscure the signal of microbial interactions.** ISME 2019 (different simulations):

*“Species interactions is assumed parsable from species’ (relative) abundances. This is false when there is any spatial structure, or heterogeneous mixtures of populations. The problem is the mismatch between the spatial scales of species interactions (micrometers) and those of microbial community samples (centimetres).”*

Also: **Species abundance correlations carry limited information about microbial network interactions.**

PLOS Comp. Bio. 2022 (different simulations):

*“Asymmetric interaction types (e.g., trophic) cannot be recovered with direction, and generally induce positive correlations. Even symmetric competition tends to show up as positive correlations (although it’s a negative interaction). This might explain why empirical correlation networks have a shortage of negative correlations.”*

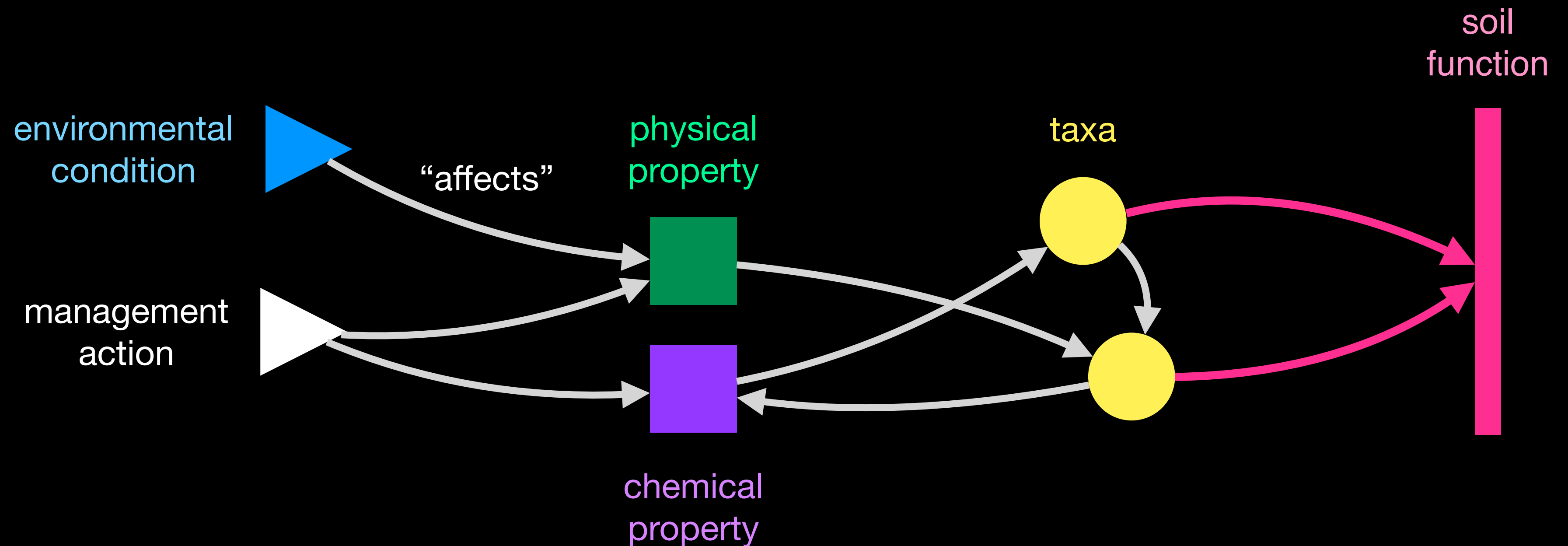
# 3.

future outlook: inferring

# functional networks we can act on

nodes: all "pillars"  
links: directed

per  
type of  
soil



biodiversity (taxa)  
abundance

physical  
analyses

soil  
taxonomy

environ.  
conditions

chemical  
analyses

manage-  
ment

function  
potential

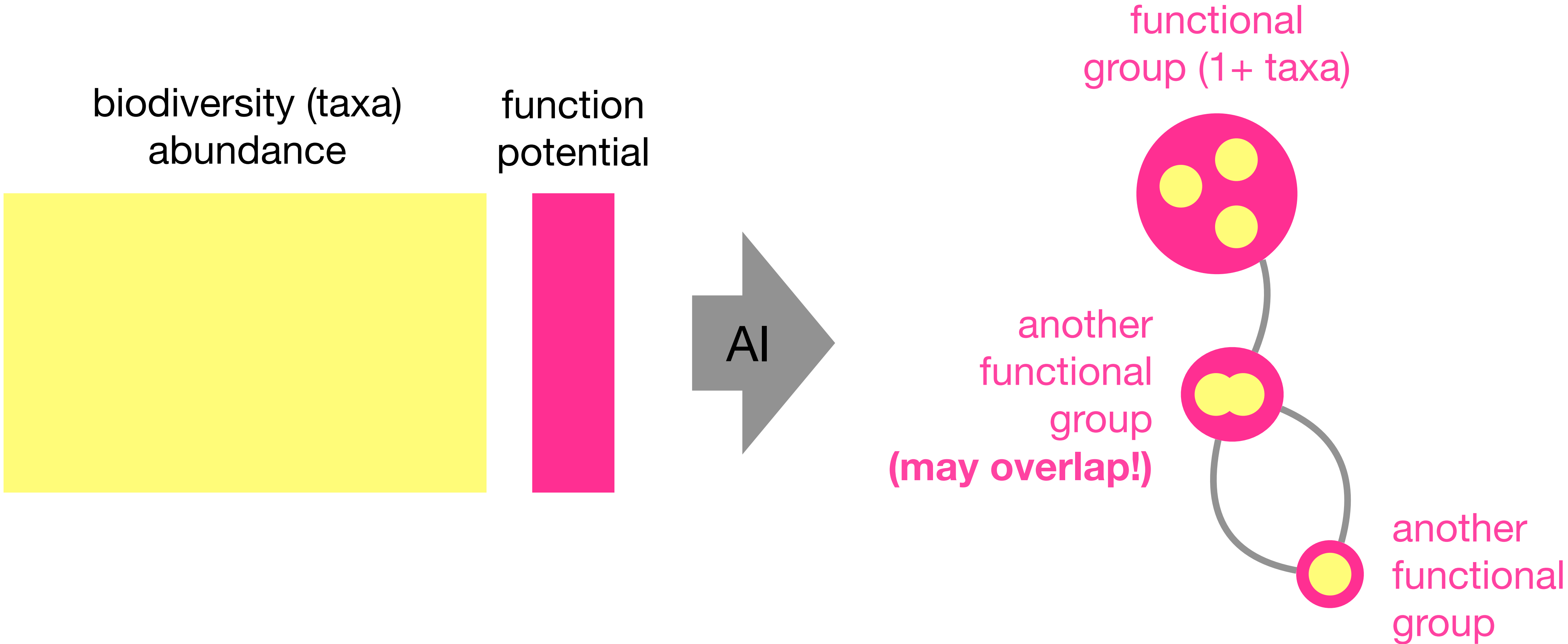


(together, a **holistic** dataset)



**predictions** of **actions** that may be undertaken to  
[fill in your problem here]

Step (1): identify functional groups of taxa, and their **function** — systematically, from data.



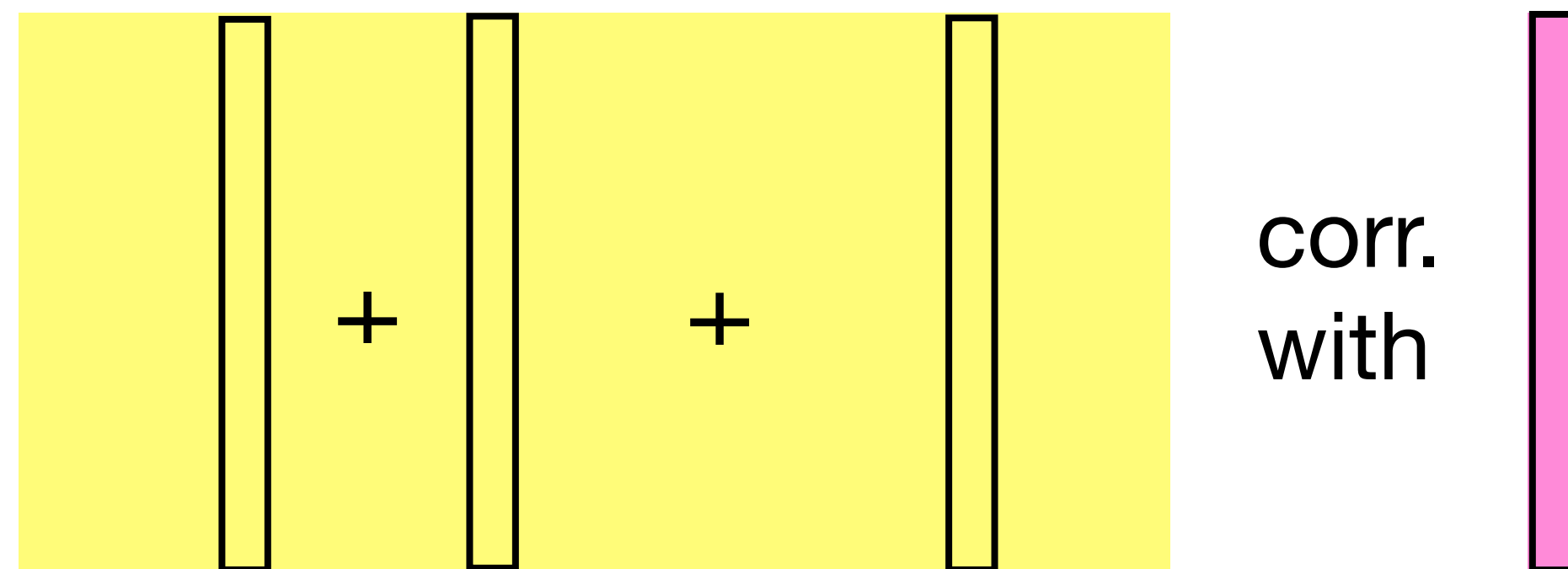
**Annotation-free discovery of functional groups in microbial communities.**

Shan et al., Nature Ecology & Evolution (2023)



**functional co-response group is:**

group of taxa whose **total abundance** best correlates with a functional variable (say, PMN)



optimise this group,  
not knowing the true group size!  
(AI = an alg. for **combinatorial optimisation: genetic algorithm**)

**gFlora: a topology-aware method to discover functional co-response groups in soil microbial communities.**

Chen, Schram, Bucur, BIOKDD (2024)



(the AI obtains better corr. if we use:)

**total “co-occurring” abundance**

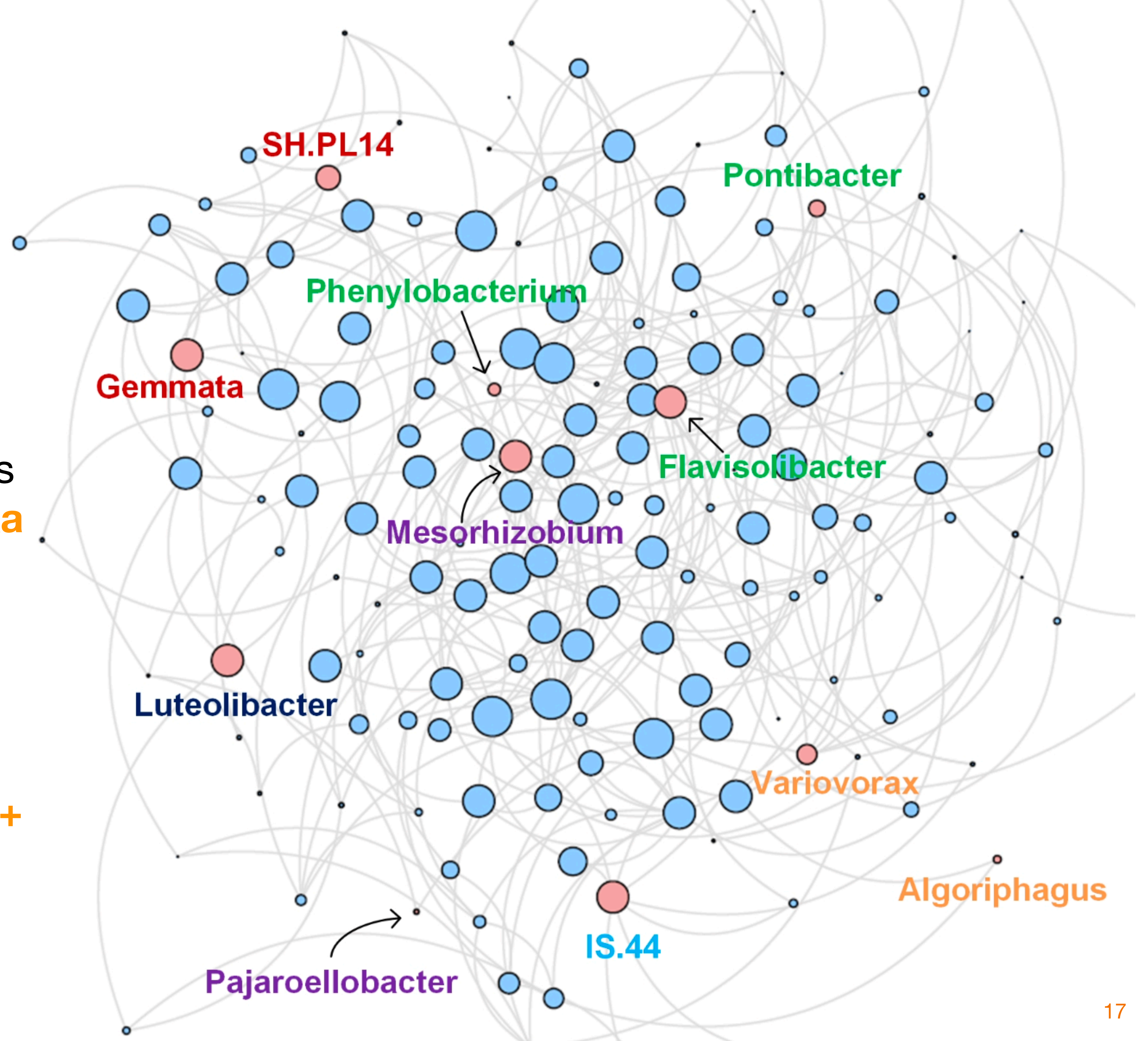


co-occurrence network  
of bacterial genera in clay,  
with the **functional group**  
corr. to PMN shown in **pink**

Our hypothesis in **gFlora**: besides  
abundance, another clue that **taxa**  
support **function**:

the taxa's **place in the co-**  
**occurrence network!**

This AI learns from **topology +**  
**abundance**.



# Future: causal “discovery” of links

(A type of AI; infers unknown causal structure from data,  
...unlike causal inference, e.g., SEM, which requires domain knowledge)

