methods for the inference of **CO-OCCUrrence** networks

(a critical assessment)

D. Bucur (UTwente, NL)

CO-OCCUrrence links from samples: close to functional?

(early results on biota)

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

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

D. Bucur (UTwente, NL)









spring



spearman







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

> On the **pairwise** assumption (more fundamentally): see

Challenge #5: Higher-order interactions (HOIs), **Open challenges for microbial network** construction and analysis, ISME 2019.

D. Bucur (UTwente, NL)

S

dense, spurious, unstable networks



(and none is entirely true)





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



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

EleMi: A Robust Method to Infer Soil **Ecological Networks with Better Community Structure.** Chen, Bucur. Int. Conf. Complex Networks (2024)

D. Bucur (UTwente, NL)

Spiec-Easi, EleMi lift the assumption of independent relationships

ground truth



neighbourhood of 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)













co-occurrence links from samples: close to functional?

Community in a Petri dish, simulation:

https://www.youtube.com/ watch?v=LwLFw1_GGnU

D. Bucur (UTwente, NL)







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)

















Bacteria

Root-feeding Nematodes

Bacterivorous Nematodes

Fungi









Can samples reconstruct all functional links? With direction?



We see *faint* groups per sample, in real abundance data (good!).

...but also *noise* (bad).

(This is: nematode species, sand, filtered, CLR-normalised abundance, purple = high.)

Species

Acrobeles mariannae Ālaimus_sp. Aphelenchoides_bicaudatus Bitylenchus dubius Clarkus_papillatus Eucephalobus oxyuroides Filenchus misellus Neopsilenchus_magnidens Paractinolaimus macrolaimus Paratrichodorus_pachydermus Pratylenchus_fallax Pratylenchus scribneri Pratylenchus_sp. Prismatolaimus_intermedius Pungentus silvestris Rotylenchus robustus Trichodorus similis Trichodorus viruliferus Tylenchorhynchus dubius Tylenchorhynchus sp. Acrobeloides_cf. Chiloplacus_propinquus Paratylenchus_projectus Pelodera_pseudoteres Pelodera strongyloides dermatitica Pelodera teres Pratylenchus_thornei Rhabditoides inermiformis Tylopharynx foetidus Aglenchus_agricola Chromadorea_X_sp. Eucephalobus_striatus Filenchus facultativus Helicotylenchus_pseudorobustus Meloidogyne_naasi Sauertylenchus_maximus Seinura demani Tylenchorhynchus teeni Aphelenchoides_sp. Aphelenchus avenae Cruznema_sp. Ditylenchus_destructor Ditylenchus_sp. Enoplea_X_sp. Meloidogyne_chitwoodi Mesorhabditis_sp. Nygolaimus_brachyuris Plectus_aquatilis Pristionchus Iheritieri Rhabditis terricola Thonus circulifer Trichodorus_primitivus Basiria_sp. Diphtherophora_obesus Distolabrellus_veechi Ditylenchus adasi Ditylenchus dipsaci Eumonhystera filiformis Filenchus sp. Heterodera sp. Irantylenchus vicinus Mylonchulus_sigmaturus Pratylenchus_crenatus -Pratylenchus_neglectus Psilenchus_hilarulus Steinernema_affine Tylencholaimus sp. Cephaloboides nidrosiensis Cephalobus persegnis Coslenchus rhombus Oscheius tipulae Steinernema feltiae Tripyla_sp. Tylenchus_arcuatus Diploscapter_sp. Drilocephalobus_sp. Filenchus chilensi Filenchus_vulgaris Lelenchus_leptosoma · Mylonchulus_brachyuris · Paratrichodorus teres Pelodera cylindrica Rhabditophanes_sp.



Samples



11

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!

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."

https://essay.utwente.nl/104377 (MSc thesis, T. vd Kuil) Modeling Complex Ecological Networks to Analyze the Impact of Soil Sampling Methodologies on Data

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





future outlook: inferring functional networks we can act on

per type of soil

environmental condition

management action



D. Bucur (UTwente, NL)

nodes: all "pillars" links: directed

soil function

chemical property







(together, a holistic dataset)

D. Bucur (UTwente, NL)

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



14

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



D. Bucur (UTwente, NL)





Annotation-free discovery of functional groups in microbial communities.

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

functional coresponse 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)

D. Bucur (UTwente, NL)

gFlora: a topology-aware method to discover functional coresponse 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

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

> the taxa's place in the cooccurrence 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)





