When relationships get complex: a general unsupervised approach for the inference on ecological interactions between species

Elina Numminen[1], **Jukka Kohonen**[1] **and Jukka Corander**[1]

[1] University of Helsinki, Finland

The dynamics of ecological communities are essentially described by the interactions the species have on each other. Such interactions give rise to interesting emergent properties of the whole system, sometimes promoting species diversity and persistence, and sometimes the opposite. Statistical inference on the nature of species interactions from observational studies is complicated by the large number of co-occurring species in a habitat, making the number of hypothesis on possible interactions even larger. Also, observational data might not contain enough observations of the desired type so that all the interesting hypothesis could be carefully assessed. We present a very general unsupervised method for studying species interactions from longitudinal observational studies on ecological communities. We assume the system state, i.e. the co-occurrence of species at certain time in a habitat, evolves as a Markov process and we cluster the state space based on the frequencies of transitions made. The obtained clusters reveal whether there exists important interactions between species, and also which of the species are interacting with whom. Therefore, our approach could serve as an important exploratory tool, for assessing firstly to what extent the data is informative on the interactions and secondly which interactions seem more evident and could be studied in more detail. The advantage of this approach is that it requires minimal amount of prior information on interactions or species demography in general. Since the method unravels the interaction network instead of the strength of each single interaction, it also suits better for systems with several species. We present our approach and the motivation to it by applying it to co-occurrence data on 5 bacterial species sharing the same habitat, human nasopharynx.

Keywords: Biostatistics, Machine learning, Epidemiology.