

Detecting higher-order co-occurrence patterns within aquatic bacterial communities across changing Nunavik permafrost landscape

Microbial ecosystems are prime indicators of health and biodiversity of northern environments. In recent years, thawing of permafrost led to the appearance of ponds and lakes in which microbial activity mediates the emission of greenhouse gases. These ponds host an unsuspected high microbial diversity and hence offer new opportunities to investigate interactions within aquatic consortia. In particular, J. Comte and colleagues sampled 38 ponds within four valleys in eastern Canadian subarctic, Nunavik, along a north–south permafrost degradation gradient and used network analysis to detect significant co-occurrence patterns. This analysis also led to the identification of “keystone species”, defined as species that co-occur with an important number of other species, indicating that they possibly play an important role in the structure of microbial communities [1].

To capture even more information about bacterial interactions in permafrost thaw ponds, we have developed a new general framework to detect co-occurrence patterns between more than two species. To do so, we have used a generalisation of the notion of networks: simplicial complexes [2]. Indeed, since networks can only encode pairwise interactions via edges, information regarding higher-order interactions, in which more than two nodes interact simultaneously, is lost. This reduction to pairwise interactions is problematic as it can lead to incomplete or wrong interpretations. The advantage of simplicial complexes is that they always include networks as substructures and encode higher-order interactions when necessary.

Our method relies on log-linear models and statistical hypothesis testing. More precisely, by iterating over all pairs of species, a network of statistically significant pairwise co-occurrences is first obtained. Then, log-linear models are fitted on all groups of three interconnected species in the network. Groups in which triple co-occurrence patterns prove to be non-random are distinguished from other groups. Non-random triples are

then combined to study quadruple co-occurrence patterns, and the procedure is iteratively applied to higher-order interactions. Moreover, for each pair, the association between the two species is quantified on a scale of -1 to 1: -1 meaning that both species tend to completely avoid each other and +1 meaning that they always appear together. Higher-order co-occurrences also benefit from these measures since they can fall into various interaction categories, such as purely positive, purely negative, or a mix of both. This provides valuable information to better identify indicator species. Indeed, one could argue that “keystone species” participating in purely positive higher-order interactions are more likely to play a central role in the ecosystem. In summary, our theoretical framework provides a classification of the interactions in aquatic bacterial communities and uncovers higher-order co-occurrence patterns otherwise undetectable under standard network analysis.

References:

- [1] Comte, J., Lovejoy, C., Crevecoeur, S., & Vincent, W. F. (2016). Co-occurrence patterns in aquatic bacterial communities across changing permafrost landscapes. *Biogeosciences*, 13(1), 175-190.
- [2] Lambiotte, R., Rosvall, M., & Scholtes, I. (2019). From networks to optimal higher-order models of complex systems. *Nature physics*, 15, 313–320