Interpretation of microbiome network clusters as functional groups

Jakub Kuncewicz[1], Maksymilian Chmielewski[2]
1 Uniwersytet im. Adama Mickiewicza w Poznaniu Wydział Nauk Geograficznych i Geologicznych Zakład Teledetekcji Środowiskowej i Gleboznawstwa,
2 Uniwersytet im. Adama Mickiewicza w Poznaniu Wydział Biologii Zakład Fizjologii Roślin
'Metagenomy różnych środowisk', Lublin 23-24.06.2025

Modelling the microbiome with co-occurrence networks consists of a step of determining correlations in the sample set, searching the topology and assigning vertices (species) to groups. It is useful to combine the SPIEC-EASI association estimator with a signed dissimilarity transformation and cluster assignment using the modularity maximisation method. The first two components form networks with positive associations (edges) of greater weight than the negative ones. Modularity maximisation groups nodes (species), preferring to form clusters of nodes that are directly and strongly positively connected, and the edges between clusters are weakly positive or negative. These clusters can represent cooperation between the analysed groups at the study site. The analysis of cluster topology allows the formulation of biological hypotheses, especially when information on trophic roles is available.

Founded from„Study@Research" Adam Mickiewicz University IDUB grant– 155/34/UAM/0002

Cluster construction

- **Signed transformation** - changes association into dissimilarity
- As input takes association matrix r* generated by methods like SPIEC-EASI (sparse matrix) and for each element calculates their dissimilarity
- **For each pair of species matrix *d* holds data of how differently they occur**
- **Results in negative edges with bigger dissimilarity values than positive edges**

$$d_{ij} = \sqrt{0.5(1 - r^*_{ij})}$$

- **Similarity calculation** - changing dissimilarity into similarity which is the final edge weight used in network parameter calculation
- High dissimilarity results in small similarity
- **For each pair of species matrix *s* holds data of how differently they occur**
- **Edge weight of originally negative association is lower than that of positive association in result**
- Species that "evade" each other will have much smaller strength of connection than species that "like" each other

$$s_{ij} = 1 - d_{ij}$$

- **Modularity maximization** - method that select node clusters that produce highest modularity value
- Modularity (Q) - measures how well the clusters are separated from each other, using similarity matrix A, degrees k of nodes i, j and function δ that is 1 if clusters of nodes i and j are the same or 0 otherwise, resolution term γ and number of edges in network m
- Positive edges between species of the same cluster increase the value, while negative edges or weak positive inside a cluster lower the value
- **Will select clusters with weak positive and negative edges being between clusters and stronger positive edges inside clusters**
- **This combinations results in clusters that combined species that "want" to coexist and may even cooperate**

$$Q = \frac{1}{2m}\sum(A_{ij} - \gamma\frac{k_i k_j}{2m})\delta(c_i, c_j)$$
$$ij$$

Biological hypotheses for group co-occurrence

	ENA	End-Sap	Myc	Npa	Npa-Sap	Pat	Pat-End	Pat-Sap	Sap	Sap-End	Sap-Pat
Cluster 1 (F: 39.77%, S: 29)	0.96	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.0	0.0
Cluster 2 (F: 3.96%, S: 13)	0.15	0.0	0.0	0.0	0.0	0.0	0.0	0.09	0.02	0.0	0.77
Cluster 3 (F: 16.07%, S: 25)	0.79	0.0	0.0	0.0	0.01	0.0	0.0	0.1	0.13	0.0	0.0
Cluster 4 (F: 6.6%, S: 18)	0.17	0.0	0.08	0.0	0.08	0.02	0.0	0.0	0.68	0.0	0.0
Cluster 5 (F: 7.01%, S: 18)	0.62	0.0	0.0	0.0	0.0	0.12	0.22	0.0	0.06	0.0	0.0
Cluster 6 (F: 6.3%, S: 19)	0.46	0.0	0.04	0.0	0.0	0.06	0.03	0.0	0.19	0.26	0.0
Cluster 7 (F: 0.61%, S: 10)	0.53	0.0	0.13	0.0	0.0	0.22	0.0	0.11	0.04	0.0	0.0
Cluster 8 (F: 5.14%, S: 20)	0.32	0.0	0.02	0.0	0.0	0.01	0.0	0.0	0.66	0.03	0.0
Cluster 9 (F: 4.82%, S: 15)	0.1	0.0	0.0	0.0	0.0	0.01	0.0	0.27	0.63	0.01	0.0
Cluster 10 (F: 5.26%, S: 16)	0.19	0.41	0.0	0.0	0.0	0.02	0.0	0.11	0.3	0.0	0.02

1. **Cluster 2:** Dominated by Sap-Pat (opportunistic pathogens) with ENA (ecologically not assigned) and minor Pat-Sap, indicating primarily saprotrophic species with pathogenic potential.
2. **Cluster 4:** Myc fungi co-occur with Sap/Nap(non-plant-associated)-Sap, suggesting Myc utilisation of saprotroph-processed organic matter. Minor pathogens present.
3. **Cluster 5:** ENA-dominated but with strong Pat(pathogen)/Pat-End(pathogen-endophyte), suggesting either pathogenic specialisation with exploitation of pathogen-killed biomass by Sap species.
4. **Cluster 6:** Complex Myc-centered interactions with Sap/Sap-End, minor Pat/Pat-End, and dominant ENA (potentially undescribed functional groups). ENA likely representing undescribed members of other functional groups.
5. **Cluster 7:** Pat/Pat-Sap and Myc co-occurrence, with a minor Sap component. This suggests a Myc-Pat interaction, where other species utilise biomass released due to pathogen activity.
6. **Cluster 8:** Sap-dominated with minor Myc/Sap-End and ENA, indicating likely Myc-Sap nutrient exchange.
7. **Cluster 9**: Pat-Sap and Sap dominance suggests Sap fungi may utilise organic matter released by Pat-Sap fungi, potentially even benefiting from predation of pathogenic (Pat) species on the Sap group.
8. **Cluster 10**: End-Sap dominated with minor Pat/Pat-Sap/Sap/Sap-Pat, indicating competitive End-Sap-pathogen interactions with saprotrophs benefiting from released biomass.

References:
- **signed transformation**: Stijn van Dongen, & Anton J. Enright. (2012). **Metric distances derived from cosine similarity and Pearson and Spearman correlations**. arXiv:1208.3145
- **modularity maximalisation**: Clauset, A., Newman, M., & Moore, C. (2004). **Finding community structure in very large networks.** *Physical Review E, 70(6).*

Future Perspectives

The presented network approach enables descritive modeling of fungal responses to environmental changes by revealing key interaction patterns. Experimental validation through synthetic ecosystems would bridge correlation with causation, particularly for economically critical pathogen-related dynamics. The framework's scalability also potentially allows for global comparisons to identify universal microbial rules versus local adaptations.

We are grateful for the guidance and support of our supervisor, prof. UAM dr hab.Władysław Polcyn from Adam Mickiewicz University. His mentorship has been invaluable to the progress of presented research.