Netclustering analysis with the CoOPLBM completion

Context of this analysis

After performing a netclustering on the raw data, we will see if the detect structure resulting in the clustering comes from the sampling effort. To test this we will use the CoOPLBM model by Anakok et al. (2022) to complete the data.

The CoOPLBM model assumes that the observed incidence matrix R is an element-wise product of an M matrix following an LBM and an N matrix which elements follow Poisson distributions independent on M.

The model gives us the $\hat{M}$ matrix, which elements are:

- 1 if the interaction was observed
- a probability, that there should be an interaction but it wasn't observed

This *completed matrix* can be used in different manners to be fed to the colSBM model.

Threshold based completions

With the thresholds, the infered incidence matrix obtained by CoOPLBM is used to generate a completed incidence matrix by the following procedure :

$$X_{ij} = \begin{cases} 1 & \text{if the value is over the threshold} \\ 0 & \text{else} \end{cases}$$

0.5 completed threshold

Here, the completion threshold is set to 0.5.

ARI of networks clustering: 0.5 threshold vs raw data

Sample based completions

The M matrix is used to sample a new X matrix which elements are the realisation of Bernoulli distributions of probability $M_{i,j}$.

$$\mathbb{P}(X_{i,j} = 1) = M_{i,j}$$

References

Anakok, Emre, Pierre Barbillon, Colin Fontaine, and Elisa Thebault. 2022. "Disentangling the Structure of Ecological Bipartite Networks from Observation Processes." arXiv. http://arxiv.org/abs/2211.16364.