

Fidgeting with Swarm's fastidious option

Milena Königshofen, Frédéric Mahé, Micah Dunthorn

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

Department of Eukaryotic Microbiology
University of Duisburg-Essen, Germany
milena.koenigshofen@stud.uni-due.de
CIRAD, UMR LSTM, Montpellier, France



The problem: Over- and under-estimation of operational taxonomic units (OTUs)

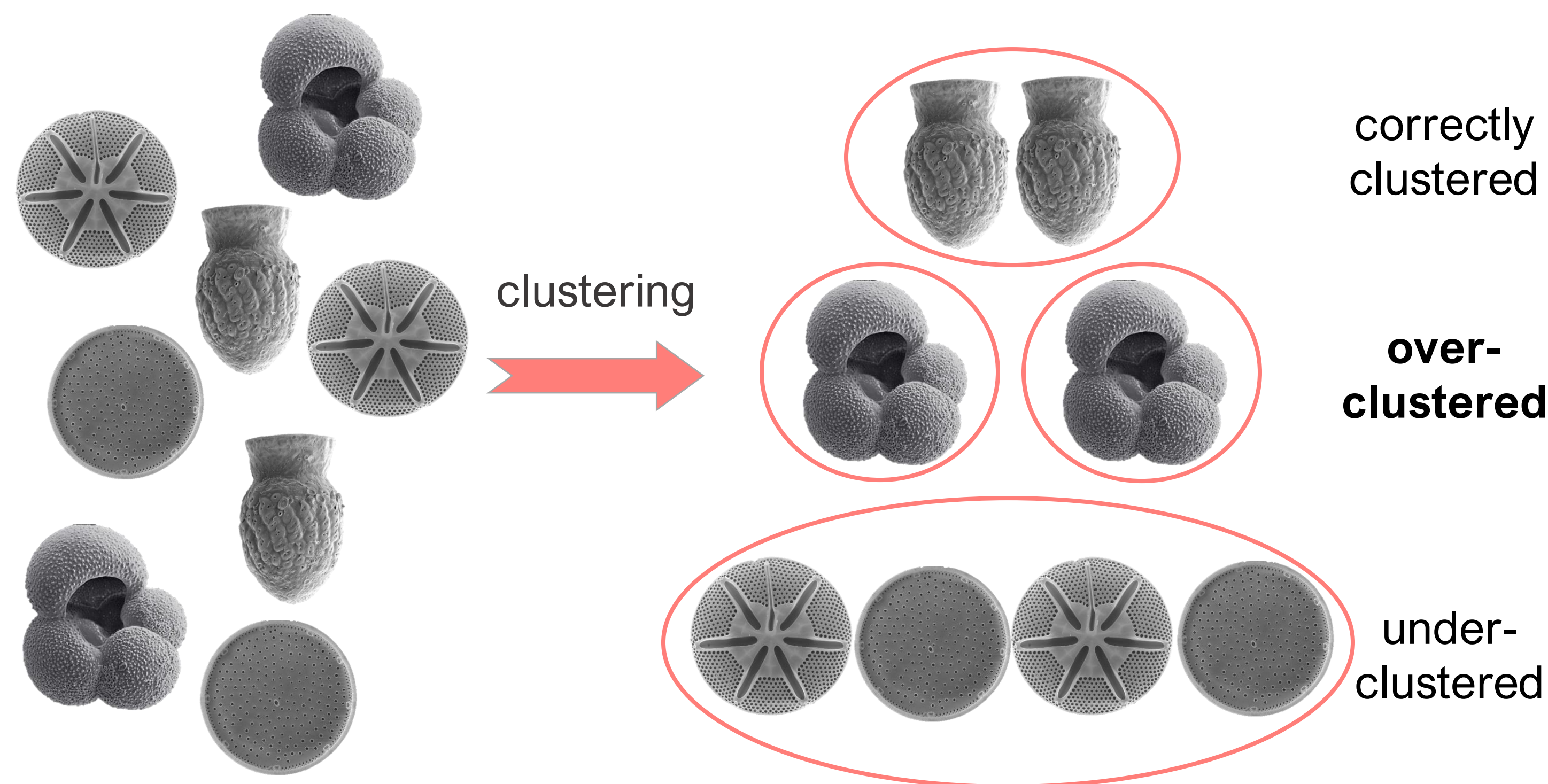


Fig. 1: OTU clustering programs can over- or under-estimate diversity, because of how they treat closely related amplicons.

How Swarm creates OTUs

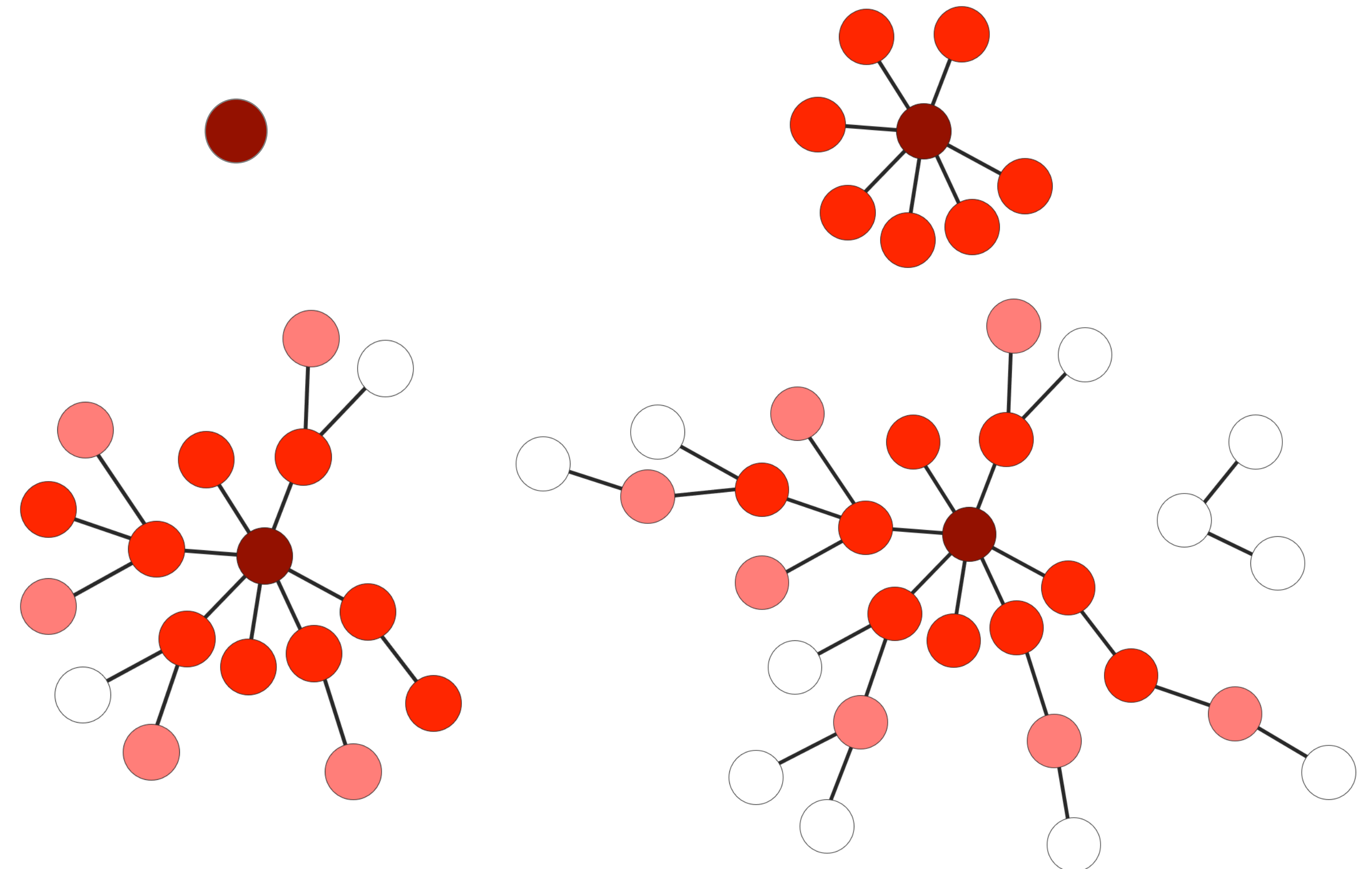


Fig. 2: Swarm (Mahé et al. 2015) clusters amplicons with a single-linkage approach that can result in over-clustered OTUs. Darker red = more abundant amplicon.

The fastidious option reduces over-clustered OTUs

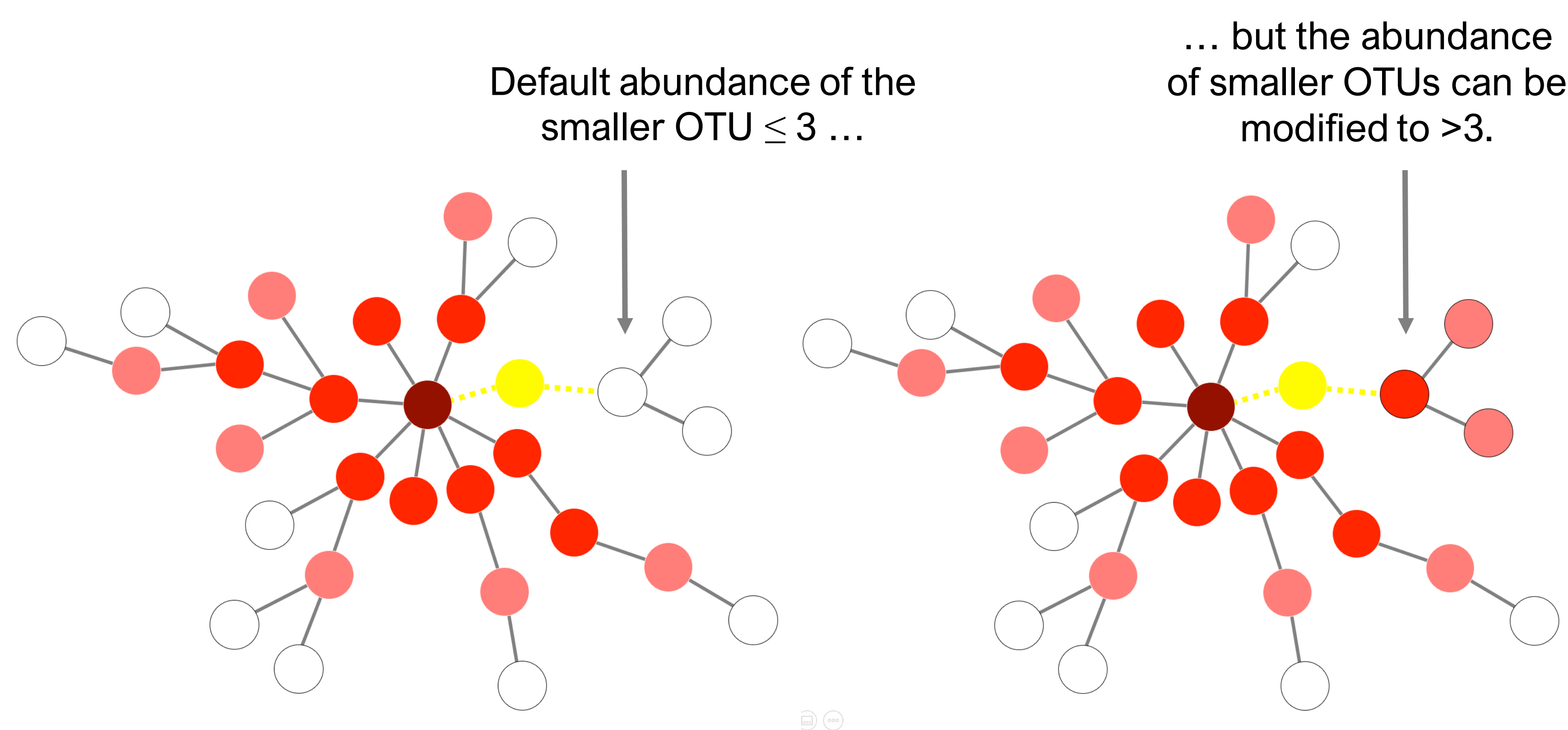


Fig. 3: Swarm's fastidious option grafts low abundant OTUs onto more abundant ones if it can postulate a linking amplicon (in yellow).

Approach to the problem

- 1) We used tropical rainforest metabarcoding data from Mahé et al. (2017).
- 2) That paper used the default fastidious option, and likely over-clustered OTUs.
- 3) OTUs were extracted if assigned to the same taxonomic reference with an identity >95%.
- 4) Amplicons in these OTUs assigned to the same reference were re-clustered with increasing fastidious option values.
- 5) We repeated this five times as a proof of concept.

Clustering results with increasing abundance values of the fastidious option

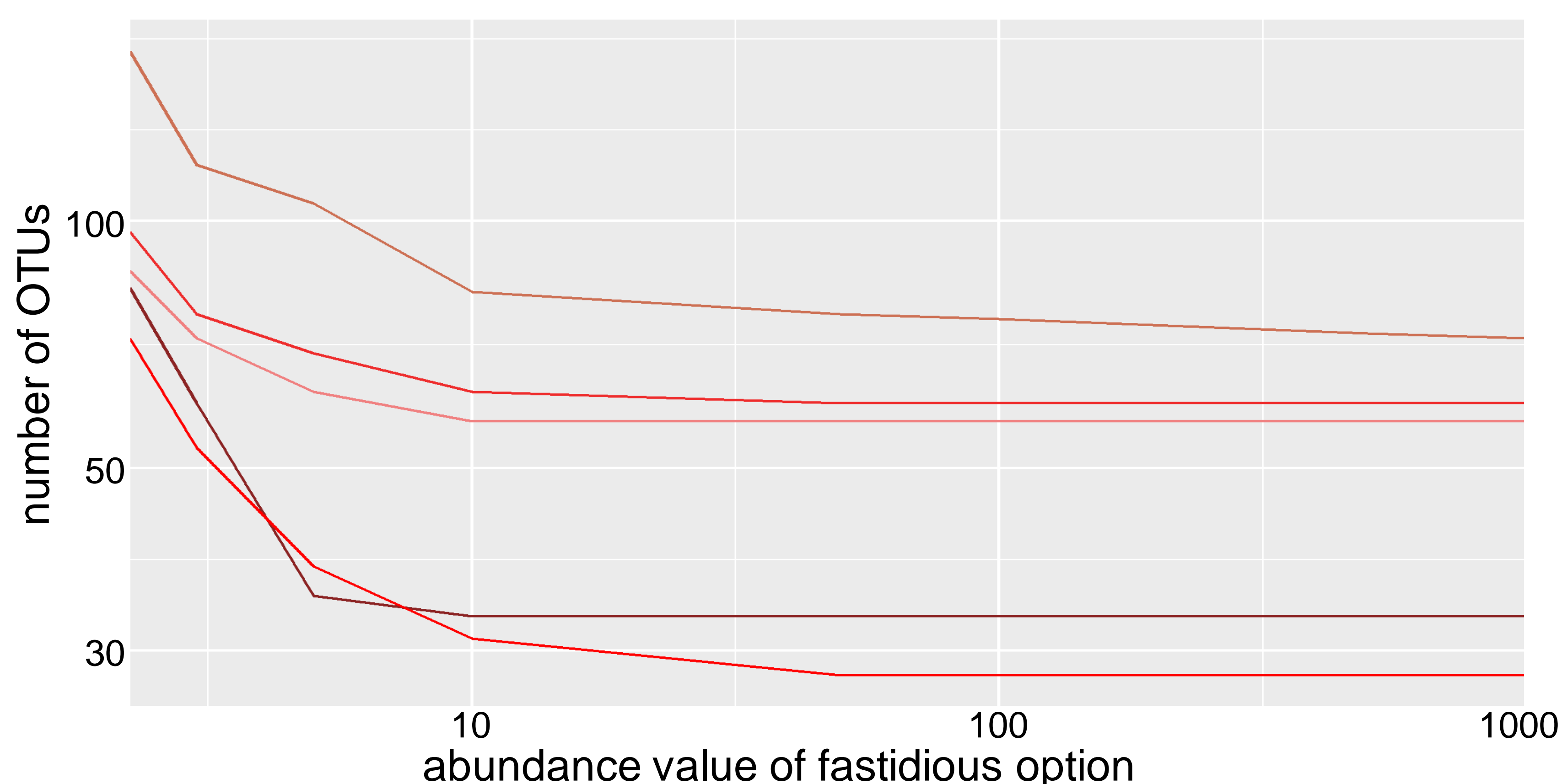


Fig. 4: The fastidious option found higher OTUs with just one missing amplicon between them and bind them at the larger OTU. Thus lowers the OTU number.

The solution: Use the powerful fastidious option.

Increasing the value of the fastidious option decreases the number of over-clustered OTUs.