Identifying stable CPD/CMTF components via low-rank graph approximation.

A step-by-step guide.

Simon Van Eyndhoven

KU Leuven

simon.vaneyndhoven@esat.kuleuven.be

Clustering: bird's eye view

1. Run an R-component tensor factorization N times
2. Construct a large (NR)×(NR) matrix with pairwise component similarity
3. Estimate a low-rank (=Q) approximation to this matrix via an EVD
4. Perform clustering in the Q-dimensional eigenspace
5. Verify which components fall within the same clusters

Cfr. accompanying slides presented at the 27th European Signal Processing Conference (EUSIPCO), A Coruña, Spain, Sep 2019

Experiment with synthetic data

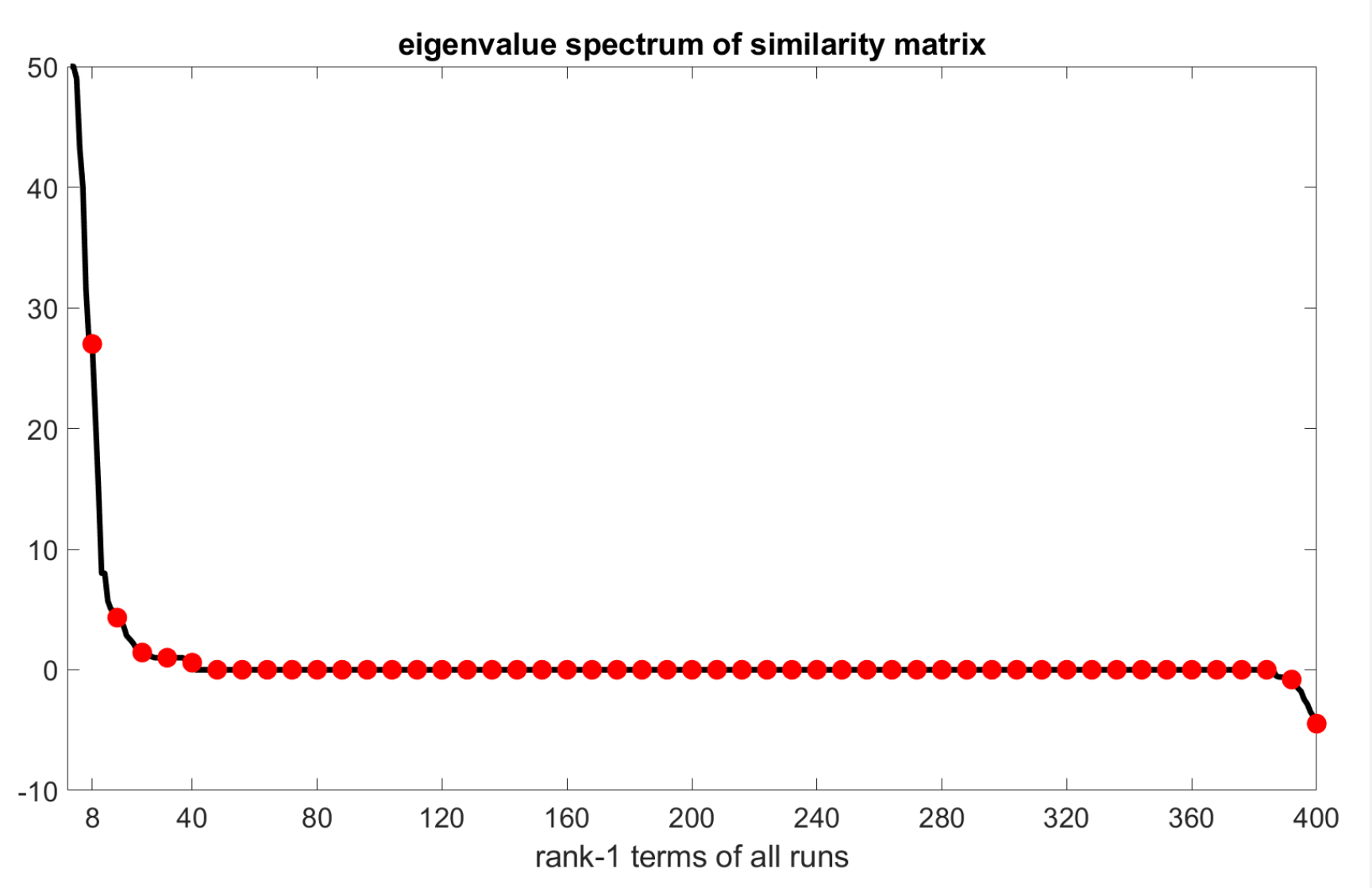
```
exp_synthcpd; % rank = 8

params.clust.Xin = 'thresh';
params.clust.maxcost = 0.05;
params.clust.roundperm = 'avglink';

[ similarity , solinfo , sols ] =
    factorizationclustering( sol , 1:getorder(T)
                           , params.clust );
```

obtain a binary similarity matrix by thresholding the continuous similarity scores

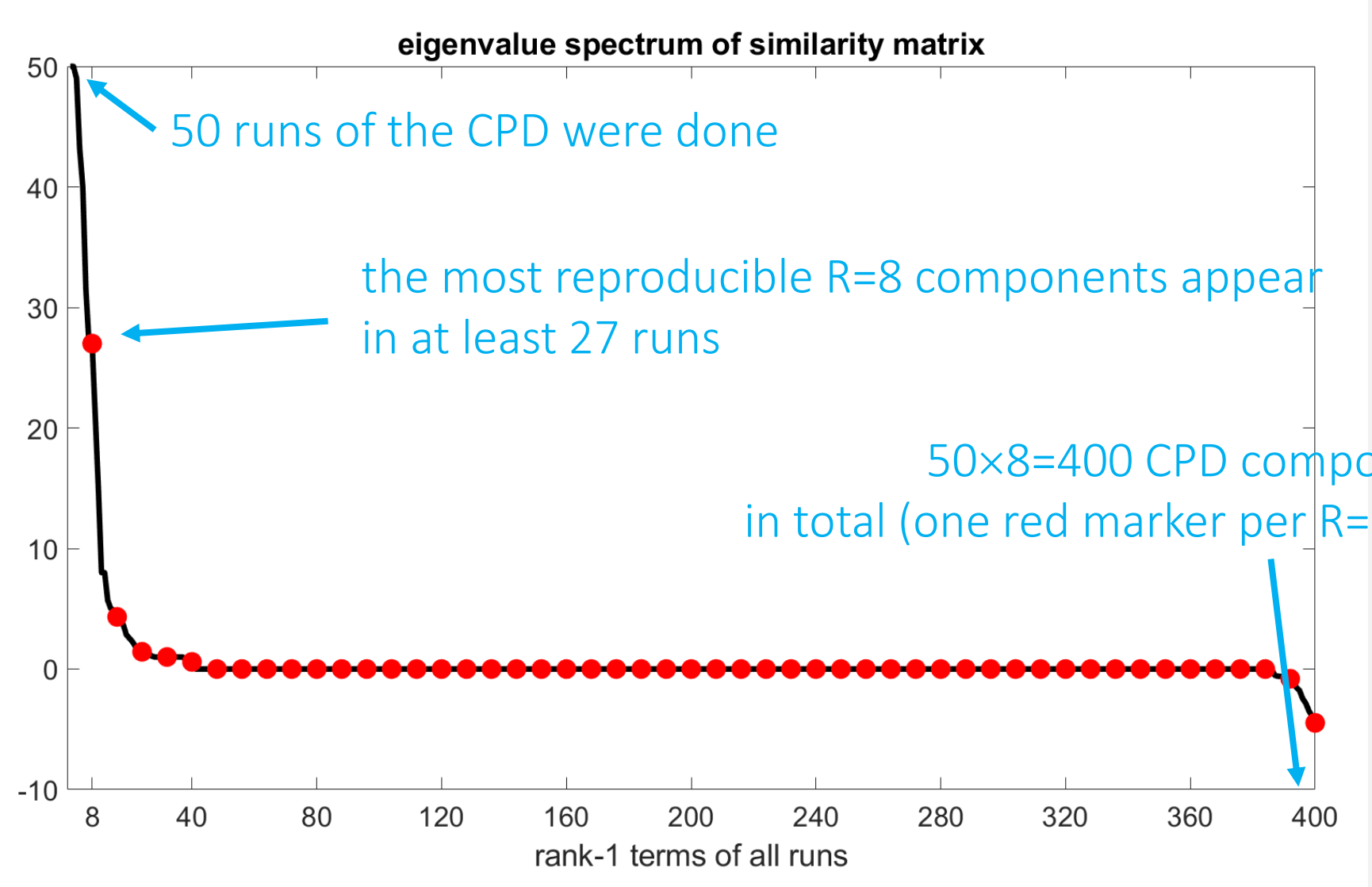
threshold = 1 – maximum ‘cost’

clustering criterion in eigenspace

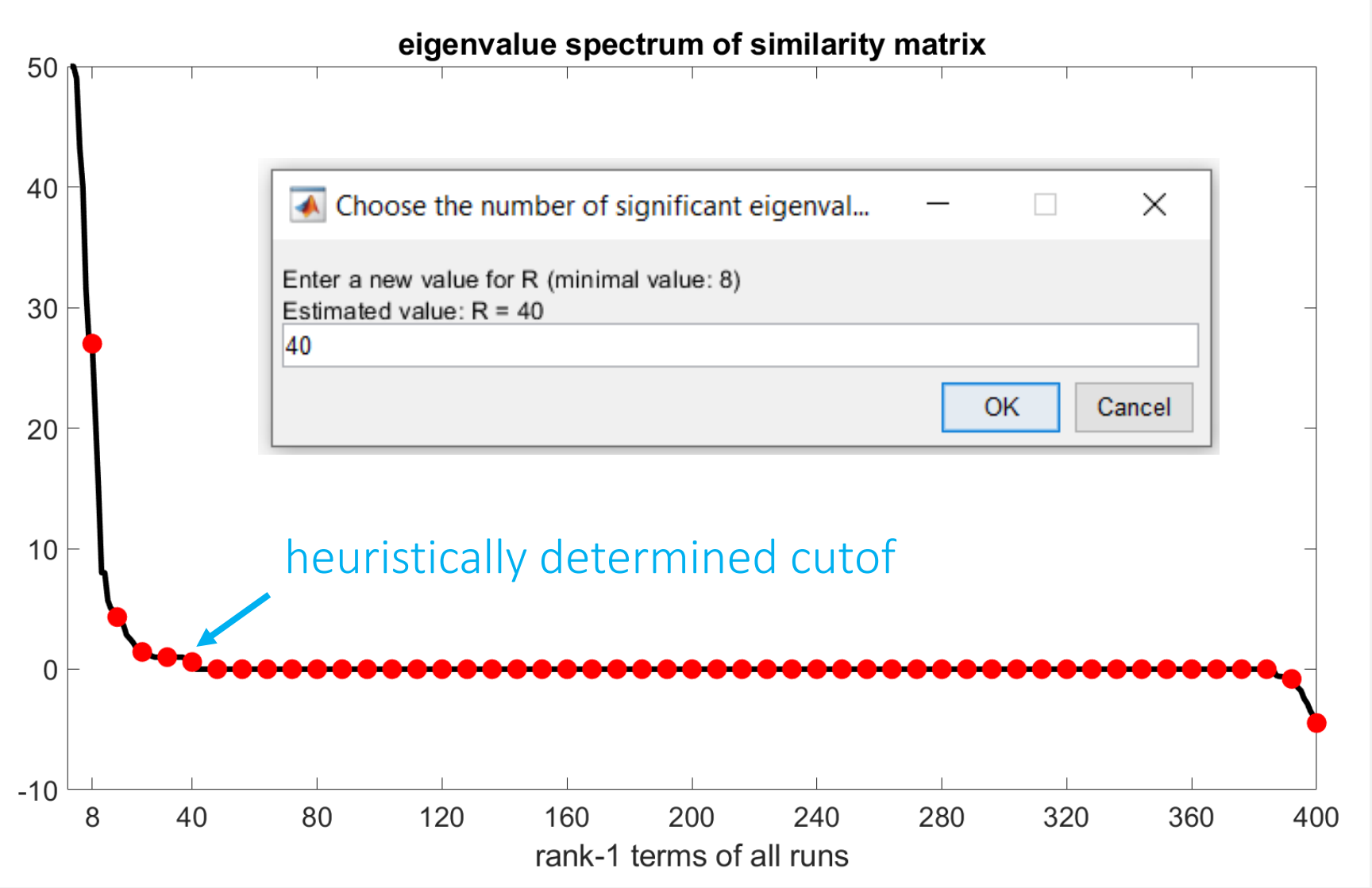
Clustering happens in a truncated eigenspace



Clustering happens in a truncated eigenspace



Clustering happens in a truncated eigenspace



Where are we at this point?

- components are individually clustered into ‘cliques’
  (i.e., each clique is a ‘club’ of very similar components from different runs)

- At a higher abstraction level, we want to know which of those cliques have members from (almost) the same set of runs
  (e.g., one might expect that in a perfect world, components appear together in groups of size R, the rank of the factorization)

- I.e., the cliques themselves can be be clustered into ‘supercliques’

- Example 1: if the factorization converges to the same rank-R solution in every run, the number of cliques is R, and the number of supercliques is 1.

Clustering the cliques

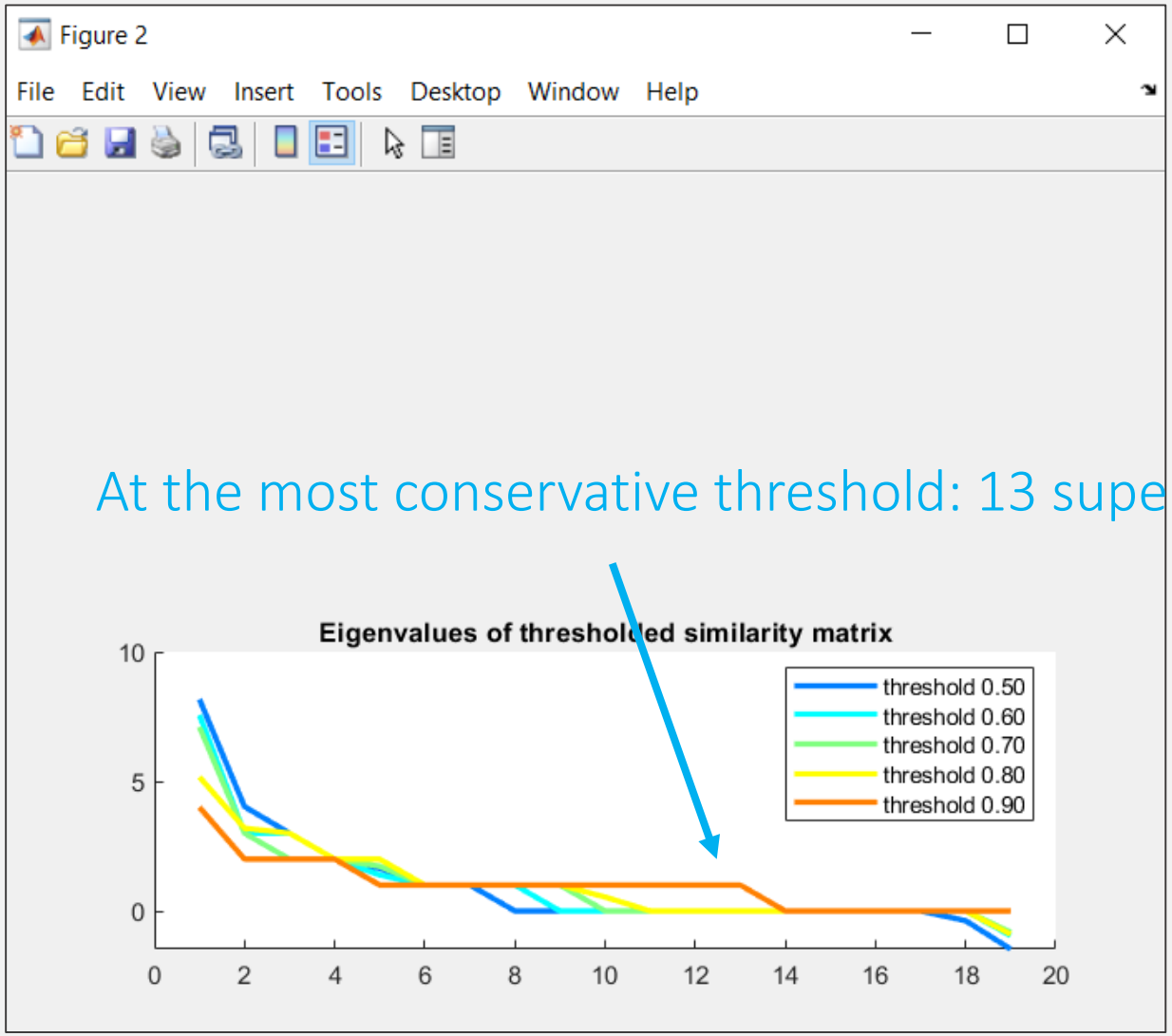


Figure 4

File Edit View Insert Tools Desktop Window Help

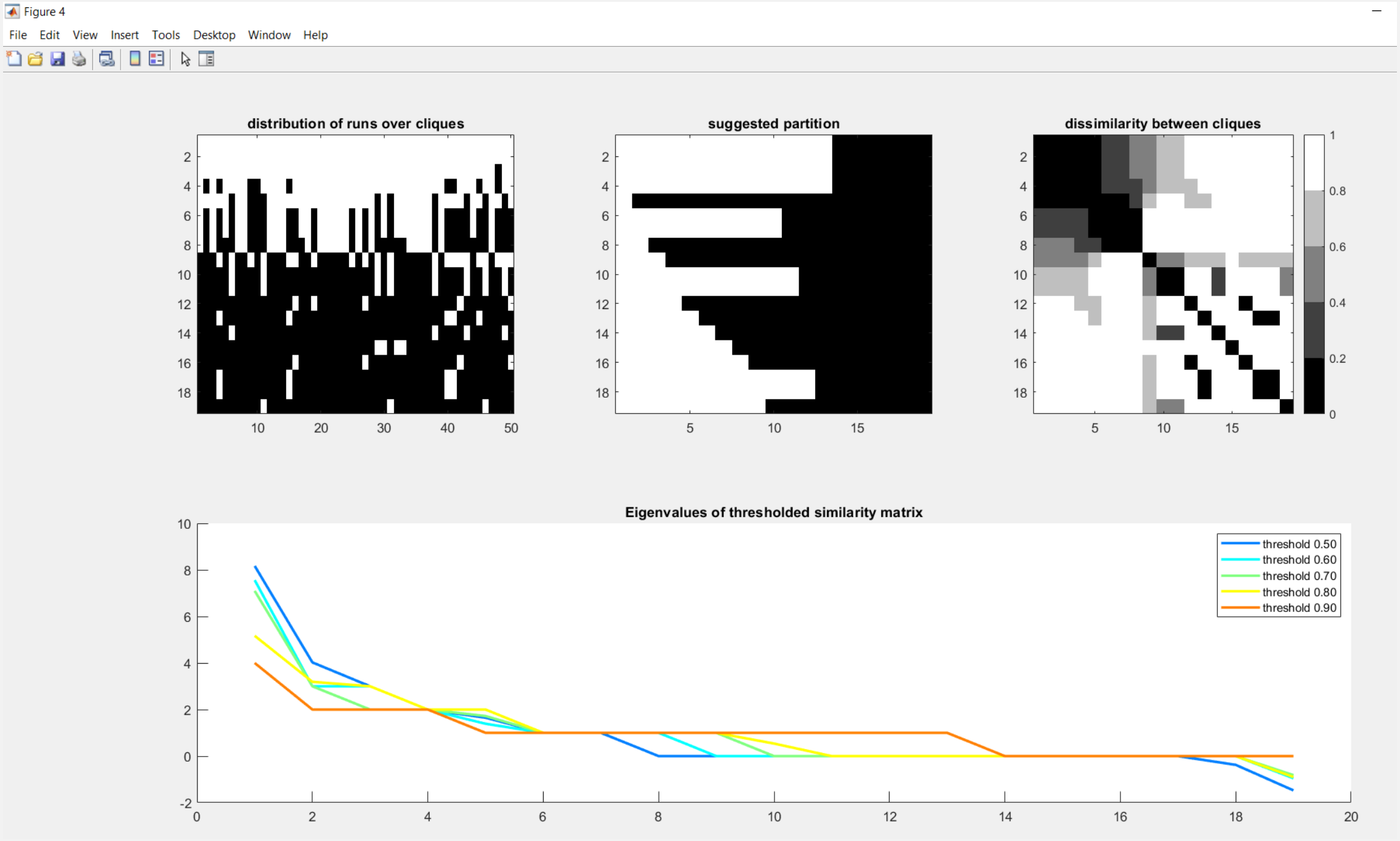
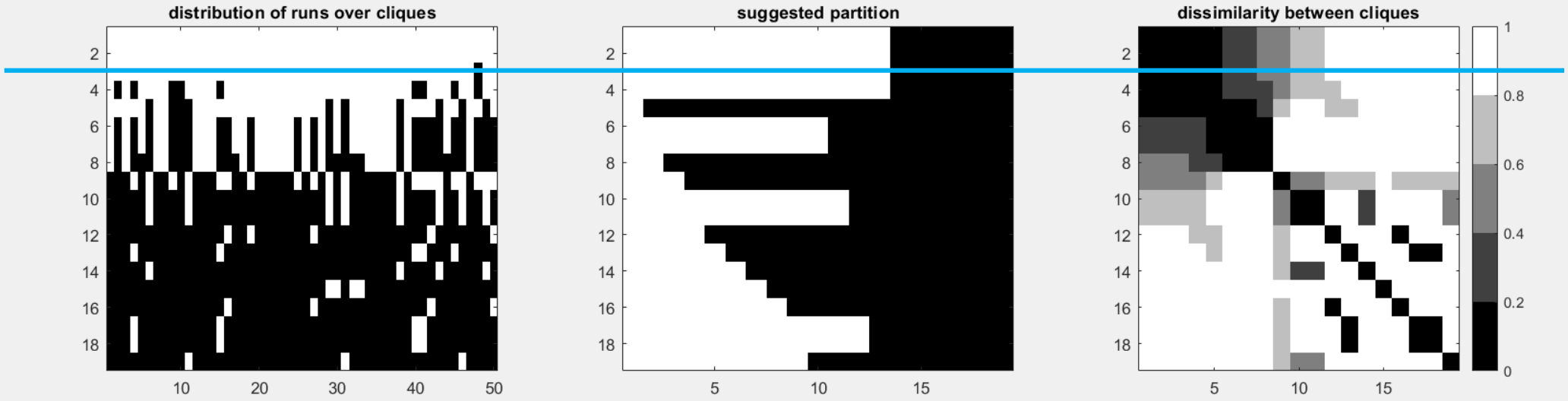
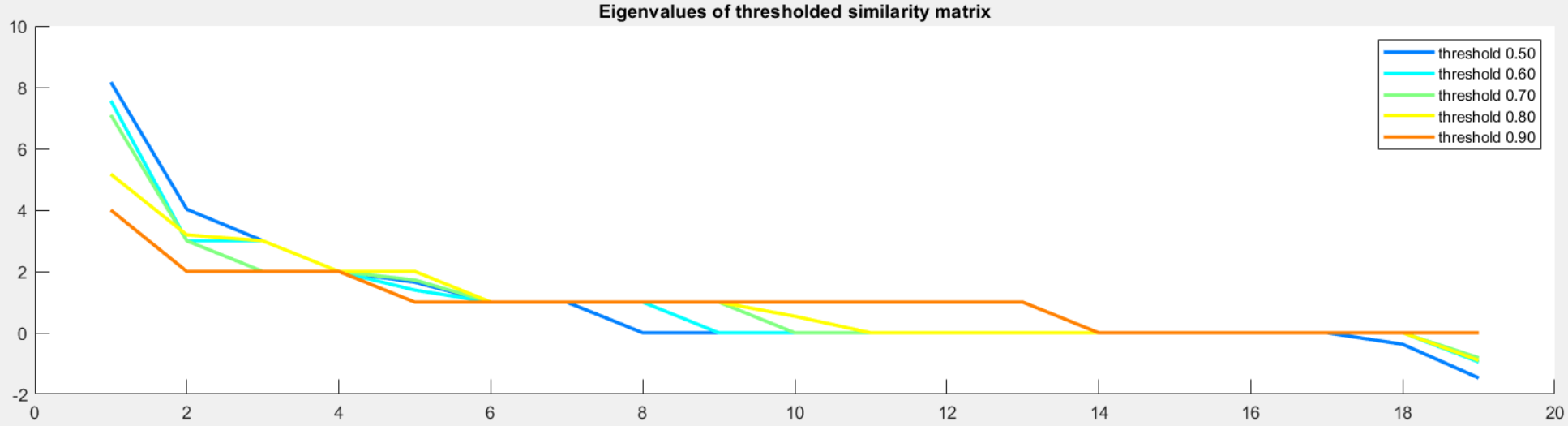


Figure 4

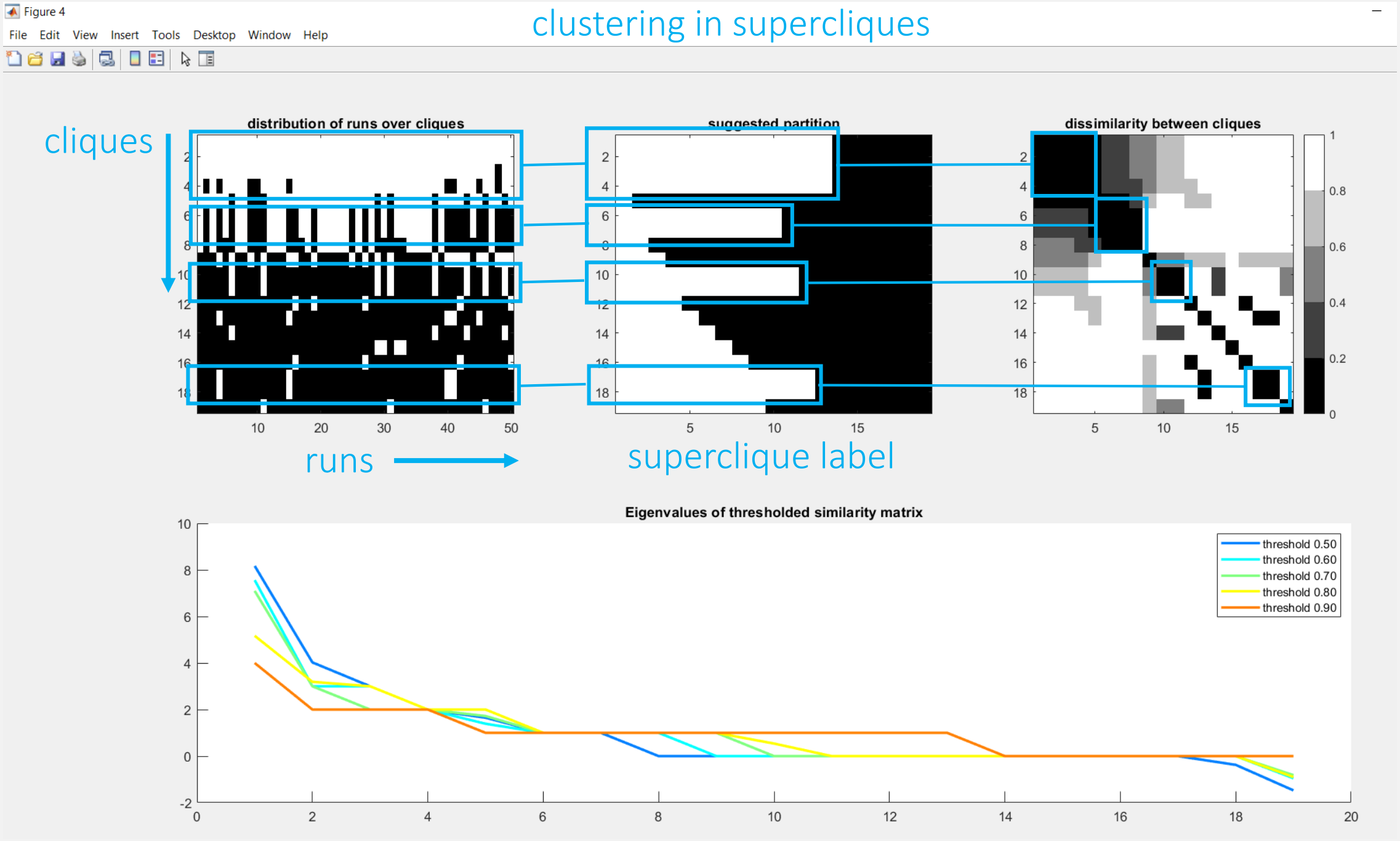
File Edit View Insert Tools Desktop Window Help

1 cross-section = 1 clique

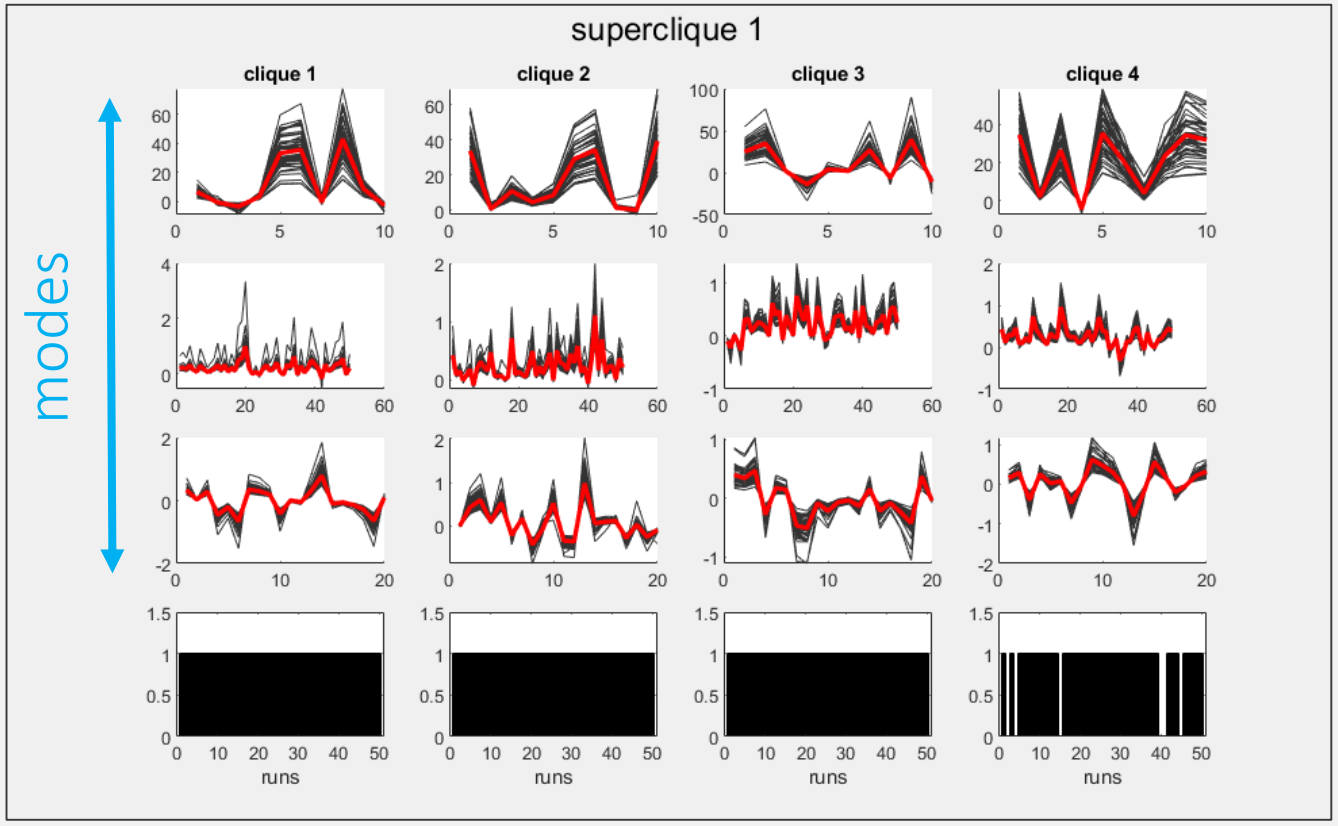


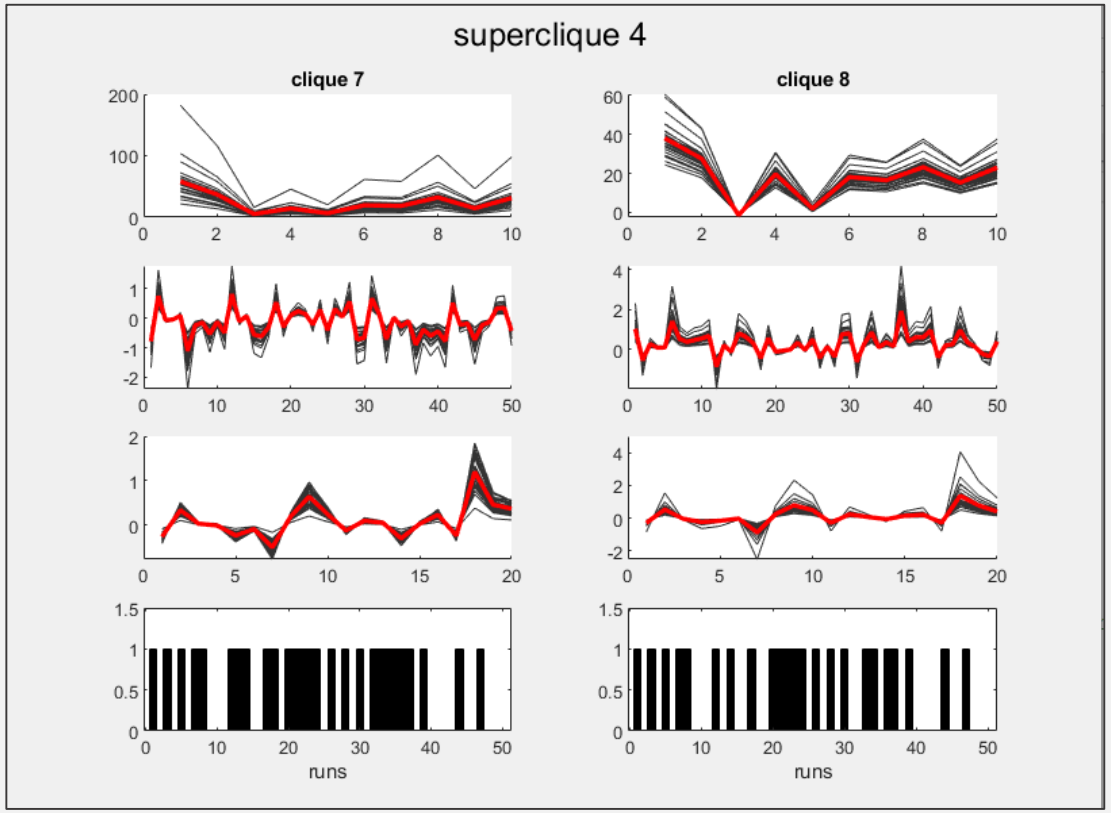


clustering in superclicques



Examples of clustering at both levels





Final notes

- Interpretation of solution quality: up to the user… ☺
- For coupled factorizations: concatenate the factors into one cell array
- SNR of the data may determine a proper similarity threshold
  very clean data: high threshold (stringent),
  low-quality data: low threshold (admissive)
- Questions? **simon.vaneyndhoven@gmail.com**