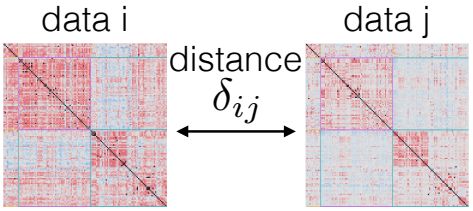
Supplemental Materials

Multi-dimensional Scaling (MDS)

- given:
  - a set of n objects
  - the dissimilarities (distances) between them δ_{ij}

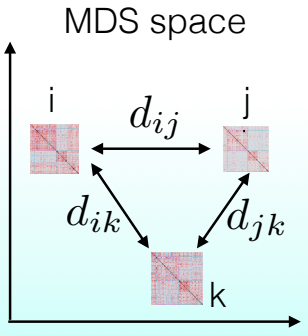


n=200

- find:
  - points on the plane whose distances d_{ij} are as close as possible to the δ_{ij}
- minimize:

$$STRESS = \left[\frac{\sum_{i,j} (d_{ij} - \delta_{ij})^2}{\sum_{i,j} \delta_{ij}^2} \right]^{1/2}$$

[kruskal 1964]

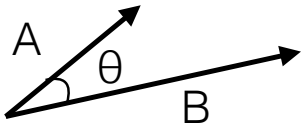


This slide based on http://robotics.stanford.edu/~rubner/slides/sld037.htm

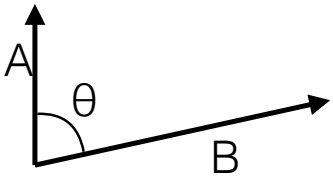
Cosine Similarity, Cosine Distance

$$sim(A, B) = \cos(\theta) = \frac{A \cdot B}{||A||\ ||B||}$$

$$dist(A, B) = 1 - sim(A, B)$$



Similar Vectors



Disimilar Vectors

- In case of Correlation Matrix

0.9	0
0	0.6

0.3	0
0	0.2

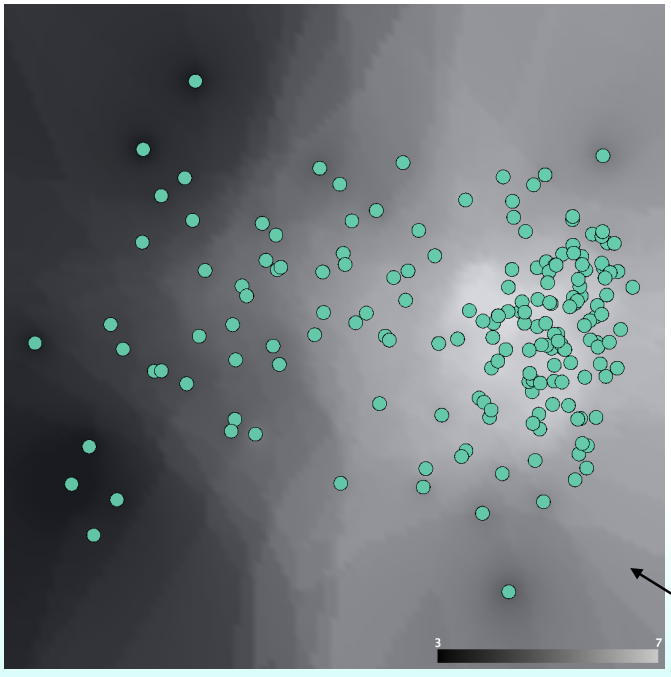
cos dist = 0
euclid dist = 0.72

0.9	0
0	0.6

0.4	0.3
0.3	0.4

cos dist = 0.22
euclid dist = 0.68

MDS Error Visualization



$$err_i = \sqrt{\sum_{j \in P} (d_{ij} - d'_{ij})^2}$$

d_{ij} : distance from point i to j in the original space

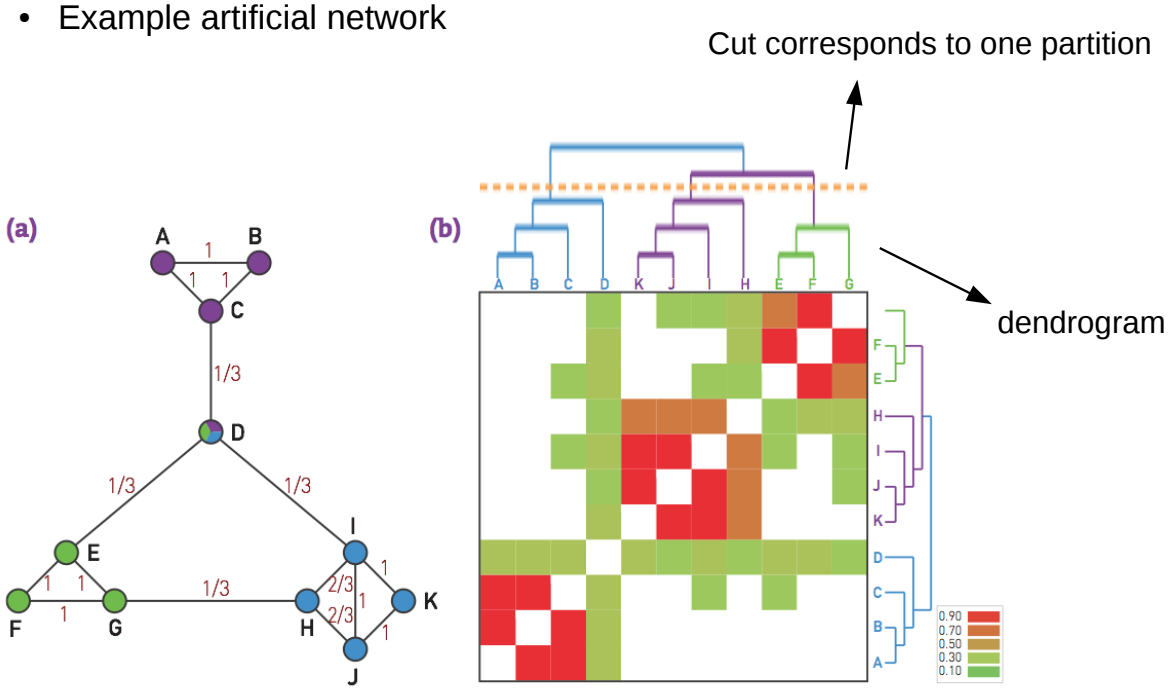
d'_{ij} : distance from point i to j in the MDS space

P : set of all points

Background color shows this error value

Hierarchical clustering: result

- Example artificial network



A.-L. Barabási, *Network Science: Communities*.
E. Ravasz et al., *Science* 297 (2002).

Modularity

- Modularity:

$$M = Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta_{s_i, s_j}$$

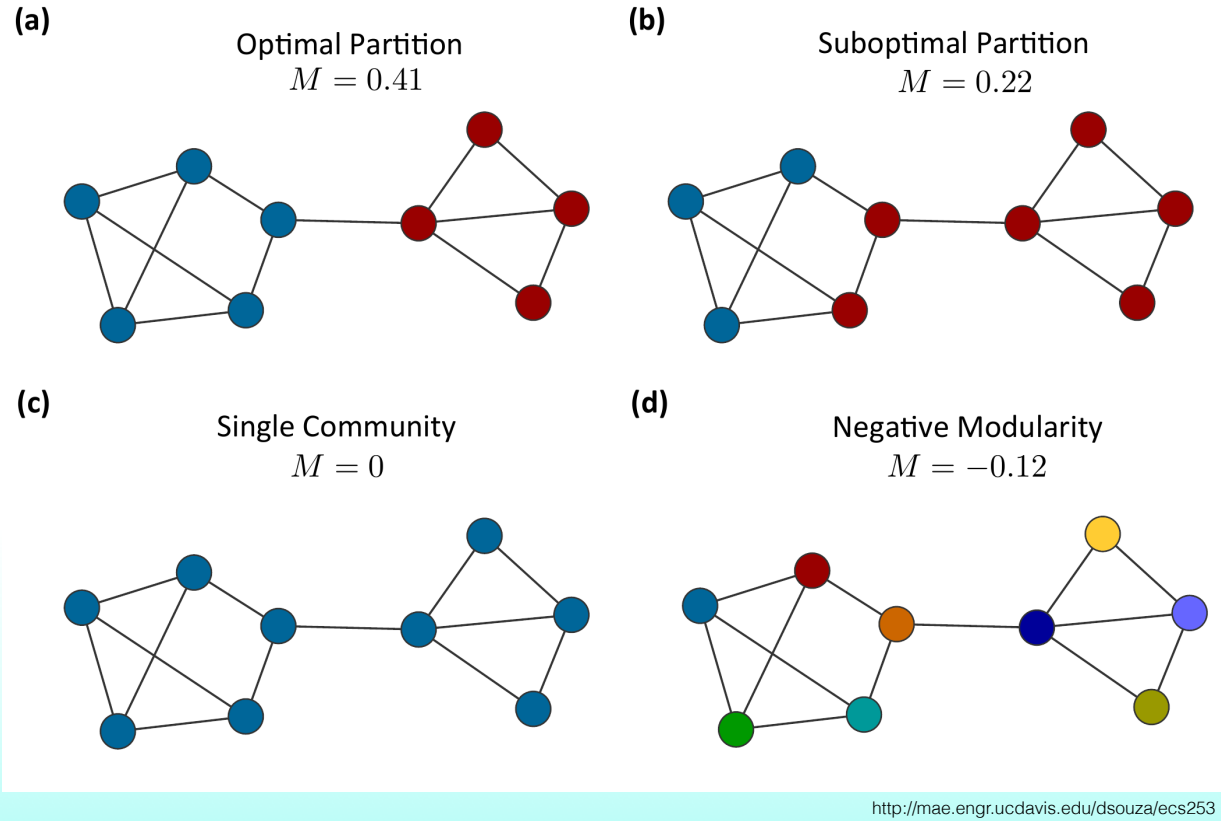
Real link

Probability of link in randomized version

- m: #links in the network
- A_{ij}: adjacency matrix, 1 if i and j are connected, 0 of not
- k_i: degree of node I
- δ_{s1s2}: 1 if in the same community, 0 if not
- High M → good division

M. Girvan et al., *PNAS*, 99:12 (2002)

Modularity



http://mae.engr.ucdavis.edu/dsouza/ecs253