Clustering Evaluation in High-Dimensional Data

Published in: M. Emre Celebi and K. Aydin, editors, Unsupervised Learning Algorithms, Springer, 2016

Nenad Tomašev[1] Miloš Radovanović[2]

[1] (Former) Artificial Intelligence Laboratory
Jožef Stefan Institute, Ljubljana, Slovenia

[2] Department of Mathematics and Informatics
Faculty of Sciences, University of Novi Sad, Serbia

Outline

- Introduction
  - Curse of dimensionality, clustering quality indexes, distance concentration, hubness
- Clustering quality indexes: an overview
  - Internal indexes
  - External indexes
- Clustering evaluation in many dimensions
  - Experimental protocol
  - Sensitivity to increasing dimensionality
    - Sensitivity of the average quality assessment
    - Stability of quality assessment
  - Influence of hubs
- Conclusion and perspectives

The Curse of Dimensionality

- The curse of dimensionality refers to different properties of high-dimensional data:
  - Sparsity (data sparsely populating the space)
  - Irrelevant features
  - “Strange” behavior of distances (distance concentration)
  - Hubness (hubs and orphans in *k*-NN graphs)
  - …
- The above are known to affect many techniques for:
  - Search and indexing
  - Classification
  - Clustering
  - …
- Effects of dimensionality on clustering **evaluation** received little attention

Clustering Quality Indexes

- Internal
  - Do not rely on outside information
  - Usually measure cluster compactness and separation between clusters, using distances (directly or indirectly)
- External
  - Based on some ground truth about the optimal partition of the data

Clustering Evaluation and Dimensionality

- One can expect internal clustering quality indexes to be affected by dimensionality
  - Distance distributions change (distance concentration)
  - Hubness appears (which indicates change in behavior of point centrality)
  - …
- Stability of indexes w.r.t. dimensionality very important when sampling feature subspaces
- We review common clustering quality indexes
  - Focus on internal
- Then, we evaluate the sensitivity (bias) and stability (variance) of clustering quality indexes with increasing data dimensionality
  - Study on synthetic data

Distance Concentration

- Ratio between a measure of spread and a measure of magnitude of distances converges to 0 as dimensionality increases

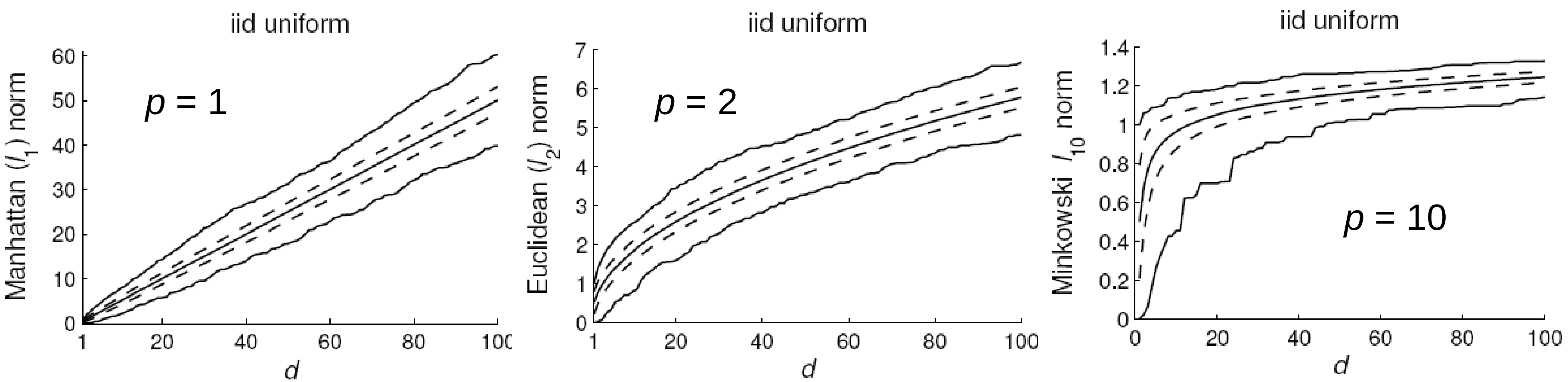
- For distance distribution *D*:
  - Relative Contrast RC(*D*) = (max(*D*) – min(*D*)) / min(*D*)
  - Relative Variance RV(*D*) = Std(*D*) / E(*D*)

- *D* can refer to distances to a particular point (conveniently 0) or pairwise distances

Distance Concentration

- **Theorem** [François, TKDE 2007]**:** For *d*-dimensional random variable $\mathbf{X}_d$ with i.i.d. components,

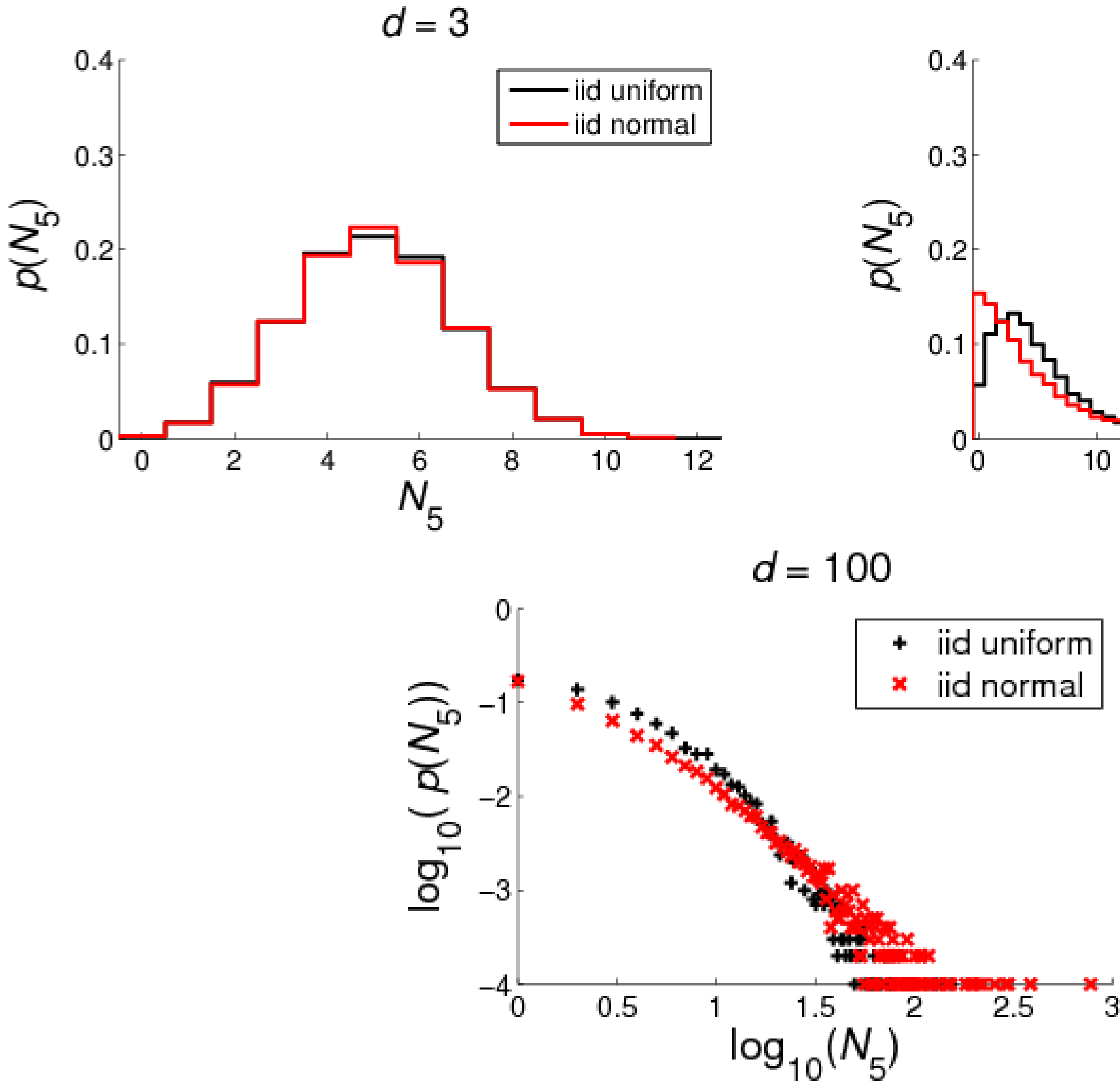
$$\lim_{d\to\infty} \frac{\sqrt{\mathrm{Var}(\|\mathbf{X}_d\|_p)}}{\mathrm{E}(\|\mathbf{X}_d\|_p)} = 0$$

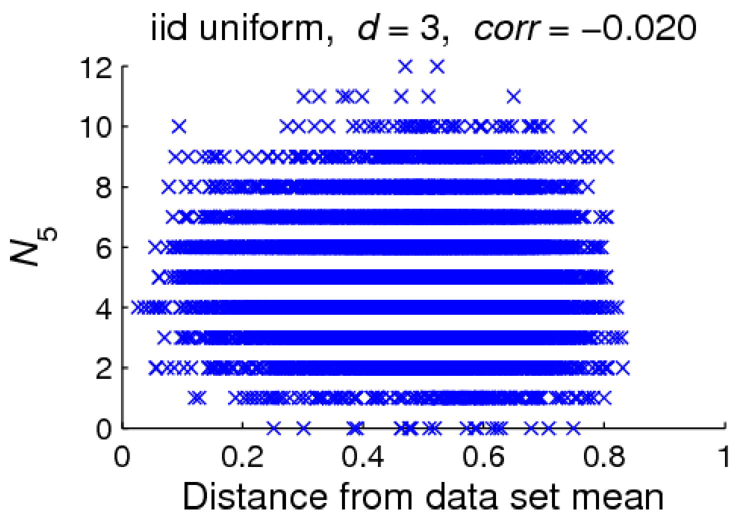


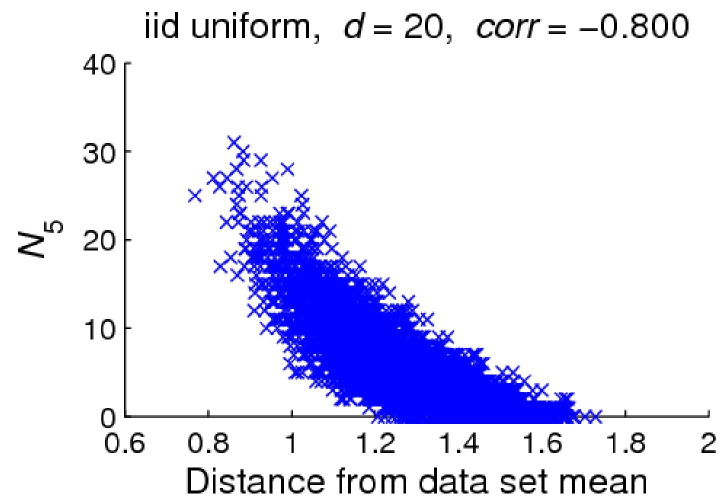
Hubness

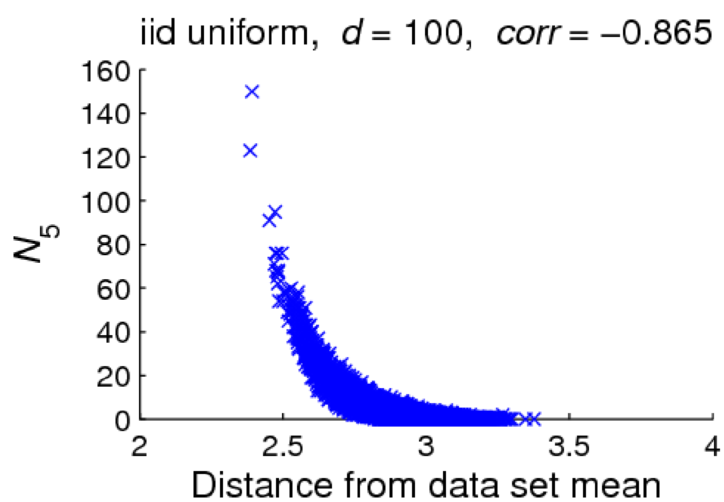
[Radovanović et al. ICML’09, Radovanović et al. JMLR’10]

- $N_k(x)$, the number of ***k*-occurrences** of point $x \in \mathbf{R}^d$, is the number of times x occurs among k nearest neighbors of all other points in a data set
  - $N_k(x)$ is the in-degree of node x in the *k*NN digraph
- Observed that the distribution of N_k can become skewed, resulting in **hubs – points with high N_k**, and **anti-hubs – points with low N_k**
  - Music retrieval [Aucouturier & Pachet PR’07]
  - Speaker verification (“Doddington zoo”) [Doddington et al. ICSLP’98]
  - Fingerprint identification [Hicklin et al. NIST’05]
  - Image retrieval [Jegou et al. CVPR’07 (talk), PAMI’10]
- Cause remained unknown, attributed to the specifics of data or algorithms


d = 3
0.4
0.3
0.2
0.1
0
p(N5)
0 2 4 6 8 10 12
N5
iid uniform
iid normal
d = 20
0.4
0.3
0.2
0.1
0
p(N5)
0 10 20 30 40 50
N5
iid uniform
iid normal
d = 100
0
−1
−2
−3
−4
log10(p(N5))
0 0.5 1 1.5 2 2.5 3
log10(N5)
iid uniform
iid normal


iid uniform, d = 3, corr = –0.020
N5
Distance from data set mean
0
0.2
0.4
0.6
0.8
1
2
4
6
8
10
12


iid uniform, d = 20, corr = –0.800
N5
Distance from data set mean
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
10
20
30
40


iid uniform, d = 100, corr = –0.865
N5
Distance from data set mean
2
2.5
3
3.5
4
0
20
40
60
80
100
120
140
160

Hubness in Real Data

- Important factors for real data
  1) **Dependent attributes**
  2) **Grouping (clustering)**

- 50 data sets
  - From well known repositories (UCI, Kent Ridge)
  - Euclidean and cosine, as appropriate

- **Conclusions** [Radovanović et al. JMLR'10]**:**
  1) Hubness depends on **intrinsic dimensionality**
  2) Hubs are in proximity of **cluster centers**

Outline

- Introduction
  - Curse of dimensionality, clustering quality indexes, distance concentration, hubness
- Clustering quality indexes: an overview
  - Internal indexes
  - External indexes
- Clustering evaluation in many dimensions
  - Experimental protocol
  - Sensitivity to increasing dimensionality
    - Sensitivity of the average quality assessment
    - Stability of quality assessment
  - Influence of hubs
- Conclusion and perspectives

Clustering Quality Indexes: An Overview

Notation:

- N – no. of data points
- $T = \{x_1, x_2, \ldots, x_N\}$ data set
- d – dimensionality
- K – no. of clusters
- $\{C_1, C_2, \ldots, C_K\}$ – partition of data set T into disjoint clusters, $\bigcup C_i = T$
- $\bar{x}$ – data-set center
- $\bar{x}_i$ – center of cluster i
- k – neighborhood size

Clustering Quality Indexes: An Overview

- Internal indexes (17)
  - Silhouette, simplified silhouette, Dunn, Davies-Bouldin, isolation, C index, C$\sqrt{K}$ index, Calinski-Harabasz, Goodman-Kruskal, G_{+} index, Hubert's Γ statistic, McClain-Rao, PBM, point-biserial, RS, SD, Tau
- External indexes (3)
  - Rand, adjusted Rand, Fowlkes-Mallows

Outline

- Introduction
  - Curse of dimensionality, clustering quality indexes, distance concentration, hubness
- Clustering quality indexes: an overview
  - Internal indexes
  - External indexes
- Clustering evaluation in many dimensions
  - Experimental protocol
  - Sensitivity to increasing dimensionality
    - Sensitivity of the average quality assessment
    - Stability of quality assessment
  - Influence of hubs
- Conclusion and perspectives

Silhouette Index

- For each point $x_p \in C_i$: [Rousseeuw 1987]

  $a_{i,p}$ – avg. distance to other points in cluster i (within cluster distance)

  $b_{i,p}$ – minimal avg. distance to other points from other clusters (between cluster distance)

$$\mathrm{SIL}(x_p) = \frac{a_{i,p} - b_{i,p}}{\max a_{i,p}, b_{i,p}}$$

$$\mathrm{SIL} = \frac{1}{N}\sum_{p=1}^{N} \mathrm{SIL}(x_p)$$

Isolation Index

[Pauwels & Frederix 1999]

- Average proportion of neighbors in the data that agree with the query point in terms of their cluster label
- Local neighborhood disagreement ratio for point p:

$$\delta_{p,k} = \frac{|x_q \in D_k(x_p) : (\nexists C_i : x_p, x_q \in C_i)|}{k}$$

- Isolation index for the data set:

$$\mathrm{IS} = \frac{1}{N} \sum_{p=1}^{N} (1 - \delta_{p,k})$$

C√*K* Index

[Ratkowsky & Lance 1978]

- Expresses contributions of individual features to within-cluster distances
- Contribution of feature *l* to the avg. overall divergence from data-set center:

$$SST_l = \sum_{p=1}^{N} \| x_p^l - \bar{x}^l \|^2$$

- Contribution of feature *l* to (inverted) within-cluster distances:

$$SSB_l = SST_l - \sum_{i=1}^{K} \sum_{x_p \in C_i} (x_p^l - \bar{x}_i{}^l)^2$$

- Final index:

$$\mathrm{C}\sqrt{\mathrm{K}}\mathrm{Ind} = \frac{1}{d \cdot \sqrt{K}} \sum_{l=1}^{d} \sqrt{\frac{SSB_l}{SST_l}}$$

Goodman-Kruskal Index

[Goodman & Kruskal 1954, Baker & Hubert 1975]

- A pair of distances is **concordant** if the distance between objects from the same cluster is lower than the distance between objects from different clusters
- A pair of distances is **discordant** if … higher …
- S_+ – no. of concordant distance pairs in the data w.r.t. the partitioning induced by the clustering
- S_- – no. of discordant distance pairs

$$\mathrm{GK} = \frac{S_+ - S_-}{S_+ + S_-}$$

G_+ Index

[Rohlf 1974]

- Takes into account only discordant distance pairs
- No. of data point pairs: $t = \frac{N(N-1)}{2}$
- Count of discordant distance pairs normalized by the total number of distance comparisons:

$$G_+ = \frac{2S_-}{t(t-1)}$$

- Lower is better, so we use the complement form:

$$\bar{G}_+ = 1 - G_+$$

Tau Index

[Rohlf 1974, Milligan 1981]

- Correlation between the distance matrix of the data and a binary matrix corresponding to whether pairs of points belong to the same cluster or not
- Can be expressed by concordance and discordance
- $t_{bw} = \binom{b_d}{2} + \binom{w_d}{2}$ – no. of distance pairs that can not be concordant or discordant since they belong to same distance type
  - b_d – no. of between-cluster pairs
  - w_d – no. of within-cluster pairs

$$\tau = \frac{S_+ - S_-}{\left(\frac{t(t-1)}{2} - t_{bw}\right)\frac{t(t-1)}{2}}$$

Outline

- Introduction
  - Curse of dimensionality, clustering quality indexes, distance concentration, hubness
- Clustering quality indexes: an overview
  - Internal indexes
  - External indexes
- Clustering evaluation in many dimensions
  - Experimental protocol
  - Sensitivity to increasing dimensionality
    - Sensitivity of the average quality assessment
    - Stability of quality assessment
  - Influence of hubs
- Conclusion and perspectives

Rand & Adjusted Rand

- No. of pairs of points: [Rand 1971]
  - a – same cluster, same label (TP)
  - b – same cluster, different labels (FP)
  - c – different cluster, same label (FN)
  - d – different cluster, different label (TN)

$$\mathrm{RAND} = \frac{a+d}{a+b+c+d}$$

- Rand prefers larger number of clusters; adjusted version [Hubert & Arabie 1985]:

$$\mathrm{ARI} = \frac{\binom{N}{2}(a+d) - [(a+b)(a+c)+(c+d)(b+d)]}{\binom{N}{2}^2 - [(a+b)(a+c)+(c+d)(b+d)]}$$

Fowlkes-Mallows Index

[Fowlkes & Mallows 1983]

- prec = TP / (TP + FP)
- recall = TP / (TP + FN)

$$\mathrm{FM} = \sqrt{\mathrm{prec} \cdot \mathrm{recall}}$$

Outline

- Introduction
  - Curse of dimensionality, clustering quality indexes, distance concentration, hubness
- Clustering quality indexes: an overview
  - Internal indexes
  - External indexes
- Clustering evaluation in many dimensions
  - Experimental protocol
  - Sensitivity to increasing dimensionality
    - Sensitivity of the average quality assessment
    - Stability of quality assessment
  - Influence of hubs
- Conclusion and perspectives

Clustering Evaluation in Many Dimensions

- Most clustering quality indexes used as
  - Objective function to be optimized
  - Criterion to make comparisons between different cluster configurations
- Assumptions:
  - Same data set (i.e. feature representation)
  - Same distance measure
- It would be useful to lift the above assumptions

Clustering Evaluation in Many Dimensions

- Clustering quality indexes are all (slightly) different, thus ensembles can be used
  - Implicit assumption: constituent indexes are equally sensitive to varying conditions in data
- For cluster configuration selection over different feature subspaces, stability w.r.t. dimensionality and representation is a strict requirement
- Our aim: shed light on sensitivity of clustering quality indexes to data dimensionality

Outline

- Introduction
  - Curse of dimensionality, clustering quality indexes, distance concentration, hubness
- Clustering quality indexes: an overview
  - Internal indexes
  - External indexes
- Clustering evaluation in many dimensions
  - Experimental protocol
  - Sensitivity to increasing dimensionality
    - Sensitivity of the average quality assessment
    - Stability of quality assessment
  - Influence of hubs
- Conclusion and perspectives

Experimental Protocol

- Synthetic intrinsically high-dimensional data sets
- Each cluster i.d. Gaussian (diagonal Cov matrix)
- No. of points: $N = 10000$
- No. of clusters: $K = 2, 3, 5, 10, 20$
- Dimensionality: d between 2 and 300
- Two settings: separated and overlapping clusters
- Generated 10 data sets for each K, d, setting
- K-means repeated 10 times
- Euclidean distance
- Clustering indexes computed on ground truth and the partitions produced by K-means

Outline

- Introduction
  - Curse of dimensionality, clustering quality indexes, distance concentration, hubness
- Clustering quality indexes: an overview
  - Internal indexes
  - External indexes
- Clustering evaluation in many dimensions
  - Experimental protocol
  - Sensitivity to increasing dimensionality
    - Sensitivity of the average quality assessment
    - Stability of quality assessment
  - Influence of hubs
- Conclusion and perspectives

Sensitivity to Increasing Dimensionality

- Synthetic data generated from same distribution type, differing only in number of dimensions
- Robust clustering quality indexes should yield similar quality scores in all cases (on average)
- Indexes sensitive to dimensionality expected to display one or both of the following:
  - Different average scores across dimensionalities – **bias** (sensitivity of the average quality assessment)
  - Large **variance** of quality predictions (instability of quality assessment)

Sensitivity of the Average Quality Assessment

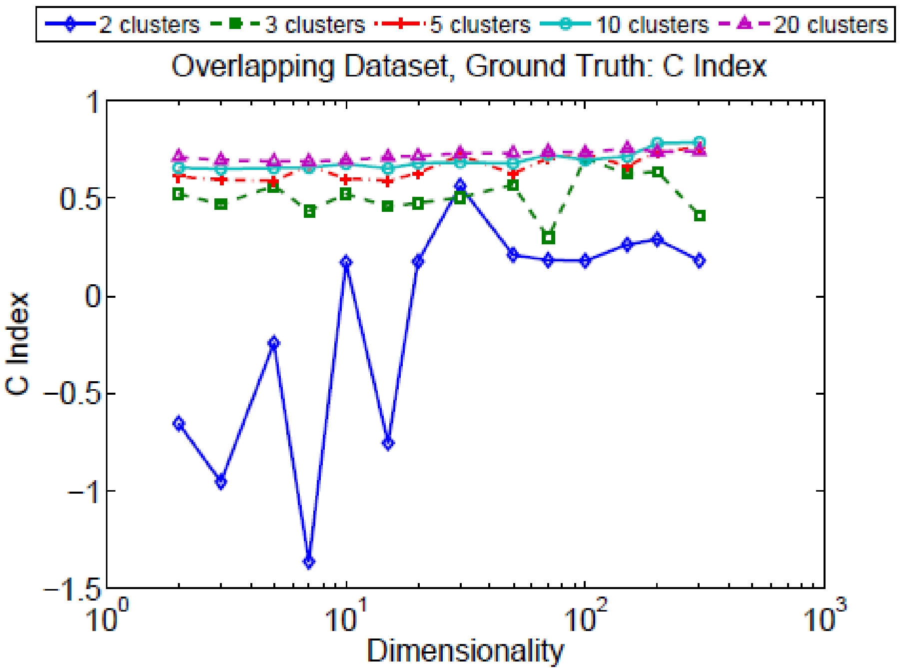
Evaluation of **ground truth**

- Some indexes seem robust to increasing dimensionality:

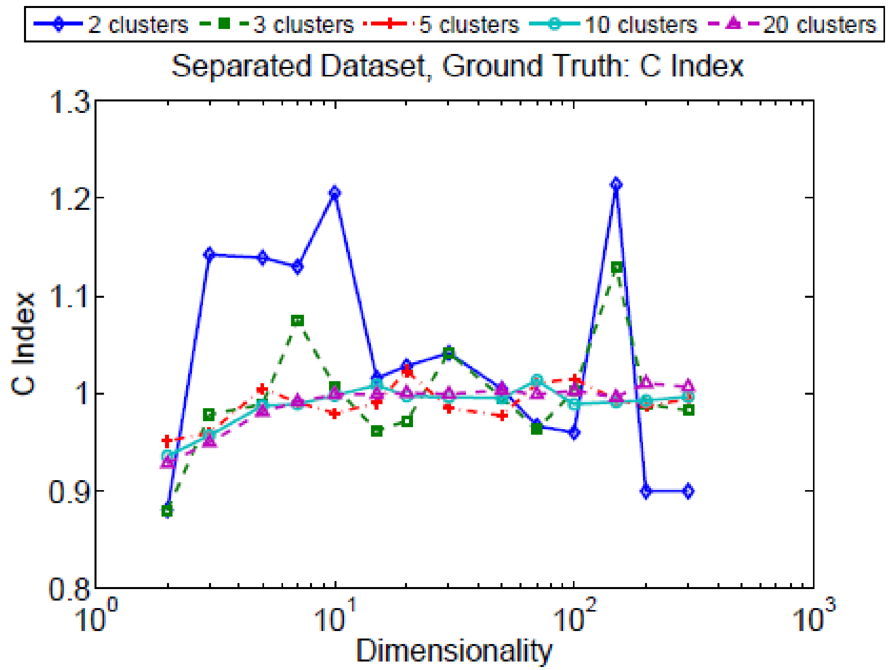
  C index, $C\sqrt{K}$ index, Calinski-Harabasz, G_+ complement, isolation, RS, Tau

- Cluster configuration quality scores remain similar when the dimensionality is increased

Sensitivity of the Average Quality Assessment: C Index on Ground Truth

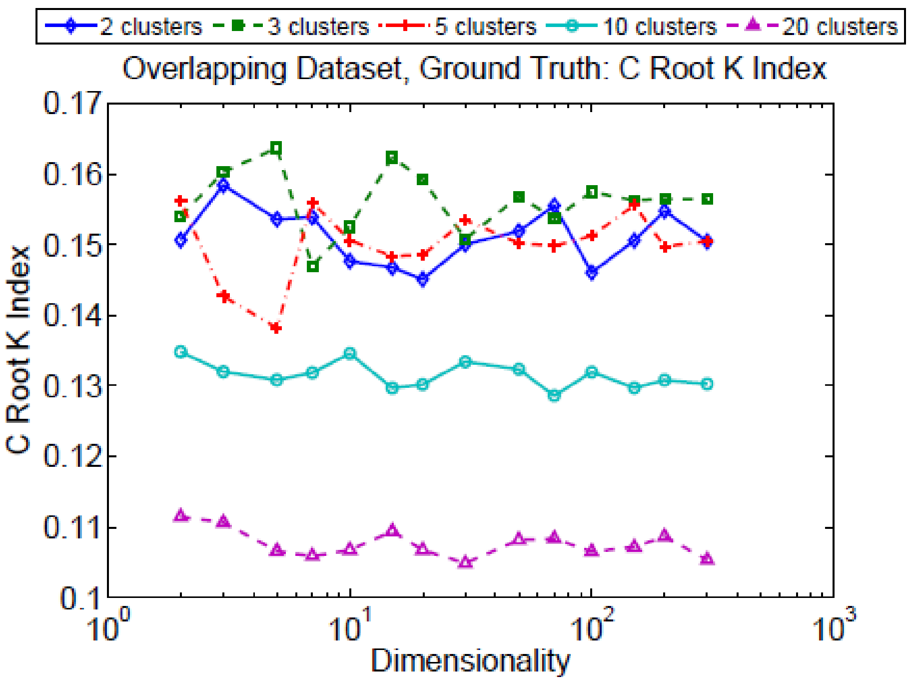


(a) Overlapping clusters

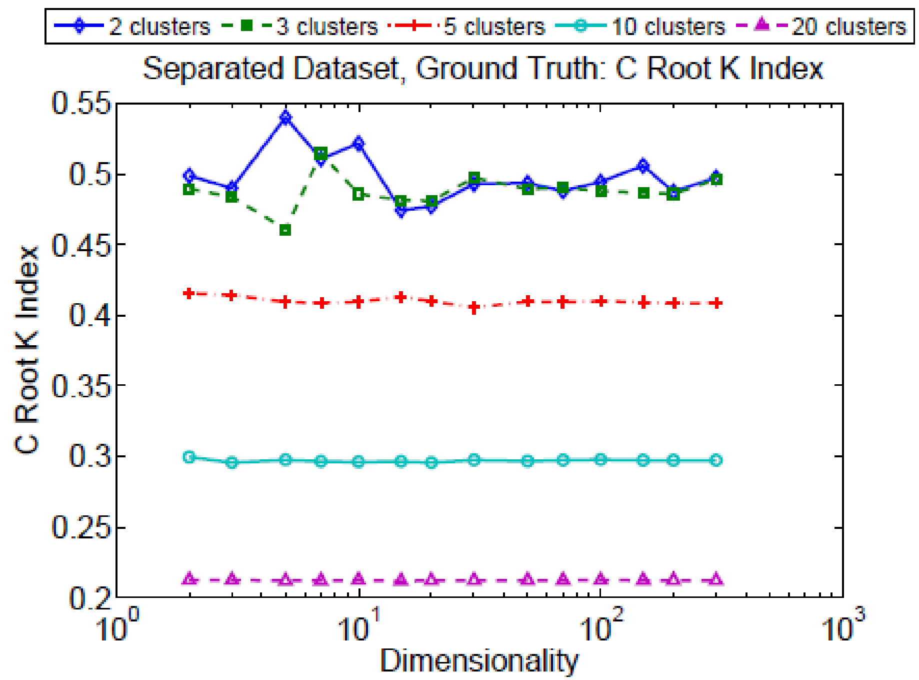


(b) Well-separated clusters

Sensitivity of the Average Quality Assessment: $C\sqrt{K}$ Index on Ground Truth

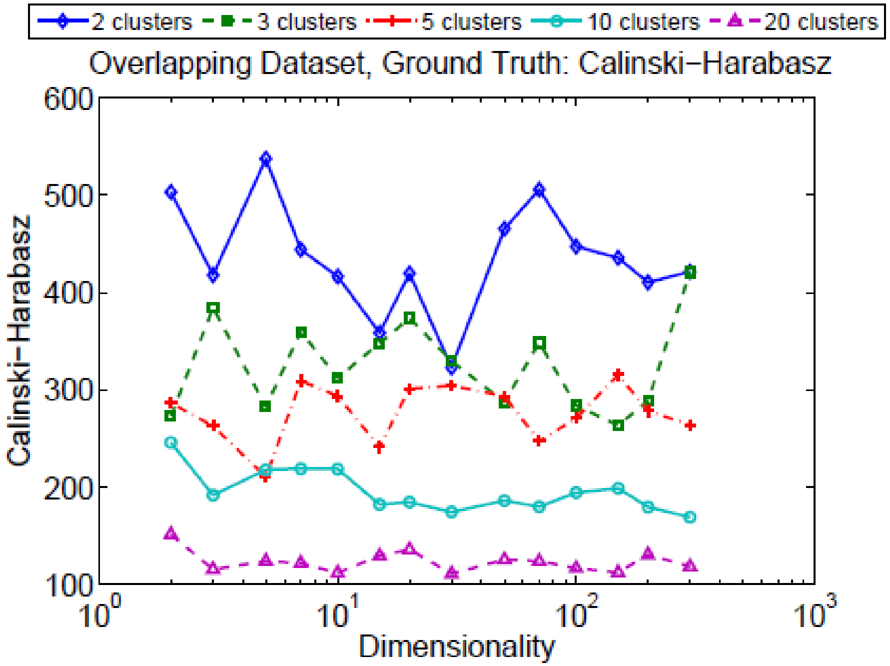


(a) Overlapping clusters

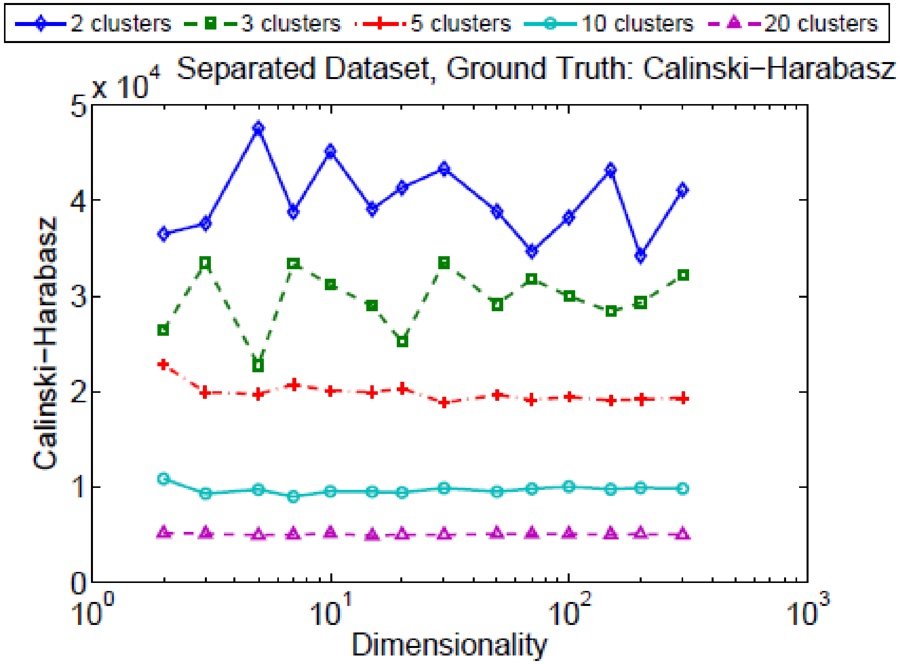


(b) Well-separated clusters

Sensitivity of the Average Quality Assessment: Calinski-Harabasz on Ground Truth

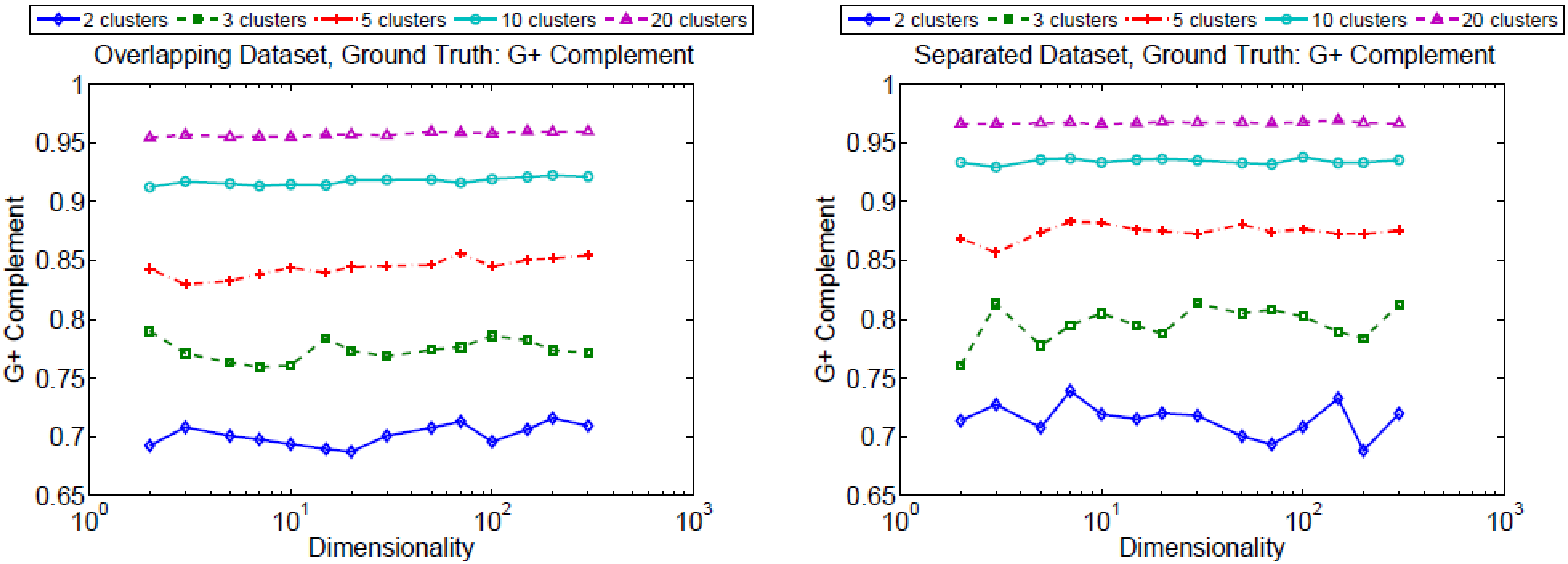


(a) Overlapping clusters



(b) Well-separated clusters

Sensitivity of the Average Quality Assessment: G_+ Complement on Ground Truth



(a) Overlapping clusters

(b) Well-separated clusters

Sensitivity of the Average Quality Assessment

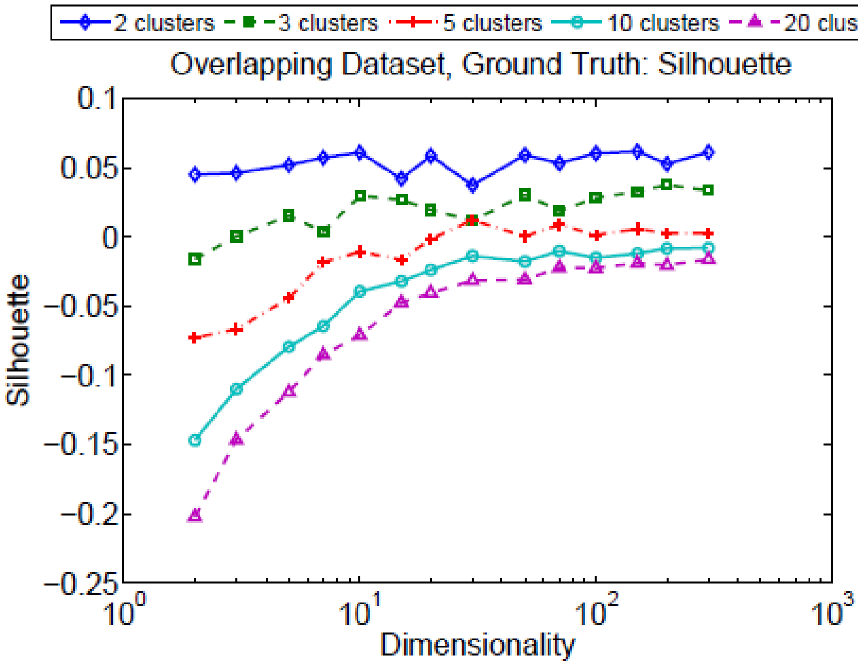
Evaluation of **ground truth**

- Other indexes are sensitive to increasing dimensionality:

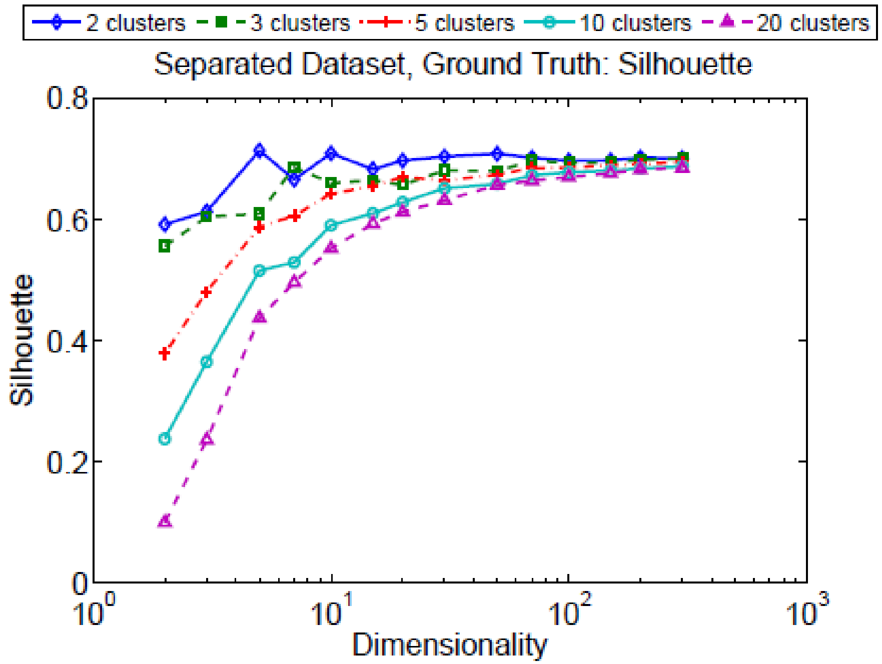
  Silhouette, simplified silhouette, Dunn, Davies-Bouldin, Hubert’s statistic, PBM, point-biserial

- Cluster configuration quality scores increase when the dimensionality is increased

Sensitivity of the Average Quality Assessment: Silhouette on Ground Truth

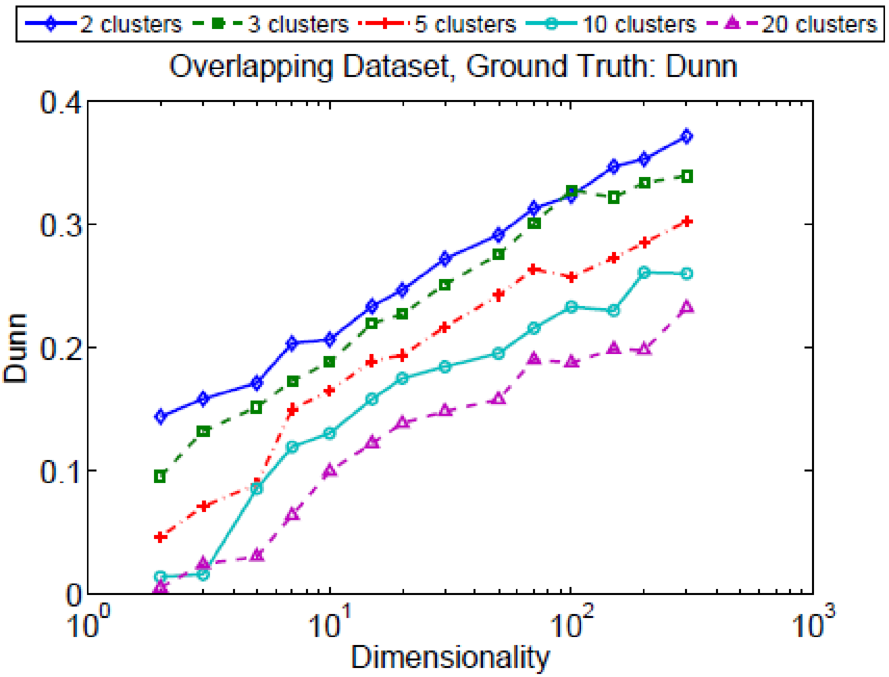


(a) Overlapping clusters

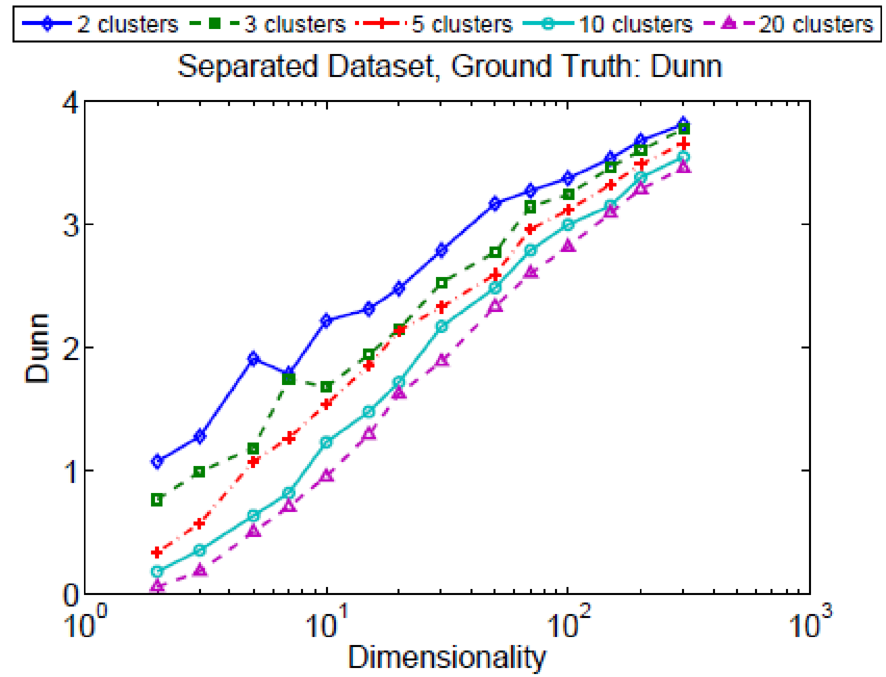


(b) Well-separated clusters

Sensitivity of the Average Quality Assessment: Dunn on Ground Truth

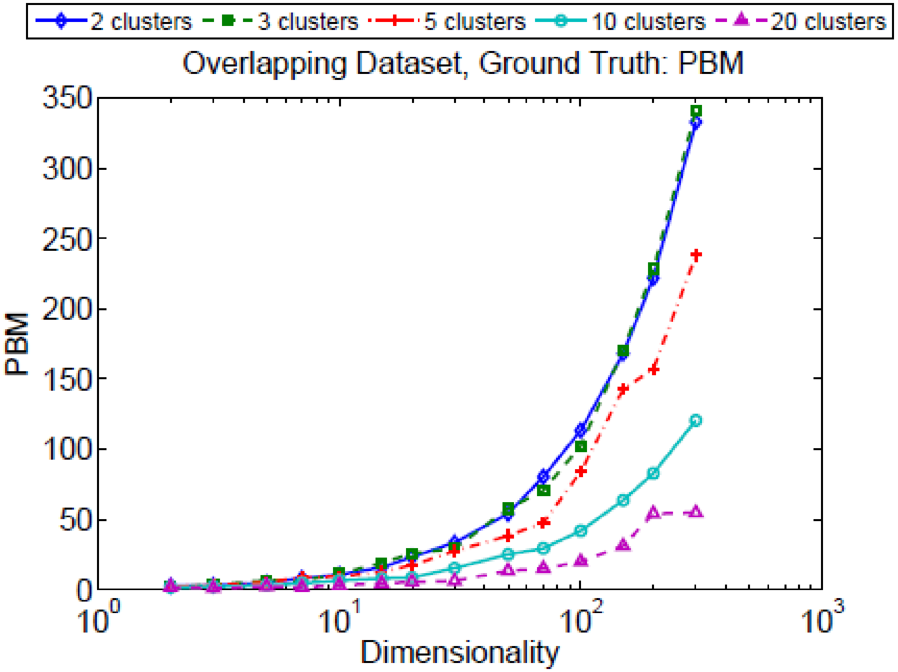


(a) Overlapping clusters

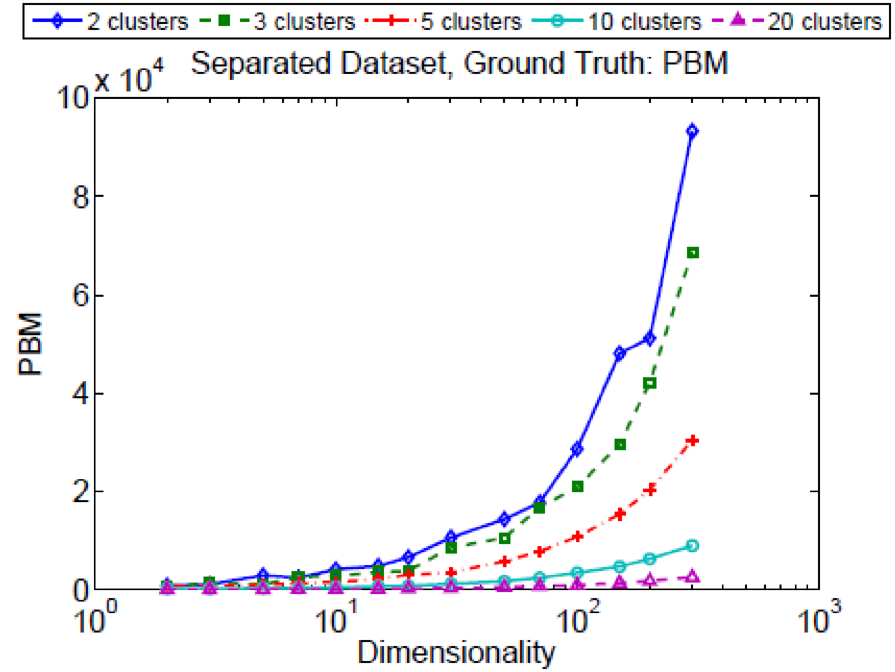


(b) Well-separated clusters

Sensitivity of the Average Quality Assessment: PBM on Ground Truth



(a) Overlapping clusters



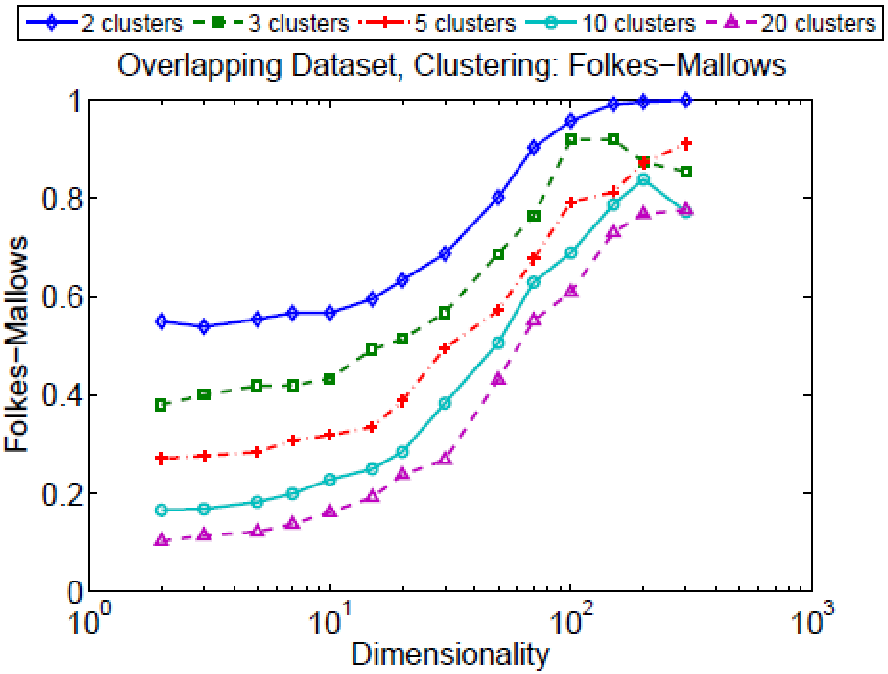
(b) Well-separated clusters

Sensitivity of the Average Quality Assessment

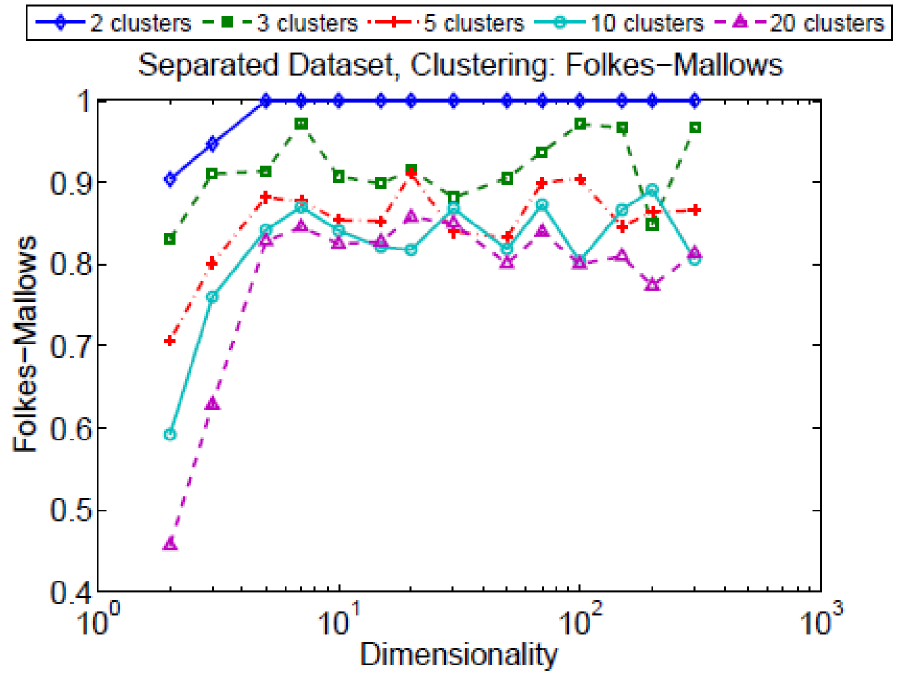
Evaluation of ***K*-means**

- Fowlkes-Mallows and adjusted Rand show that *K*-means was more successful in high dimensions w.r.t. the ground truth

Sensitivity of the Average Quality Assessment: Fowlkes-Mallows on *K*-Means

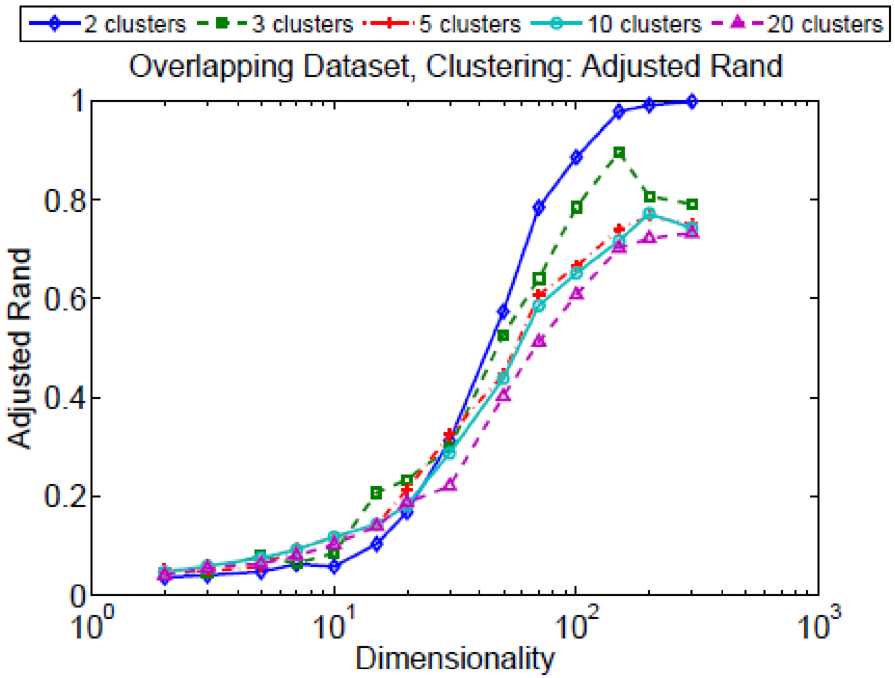


(a) Overlapping clusters

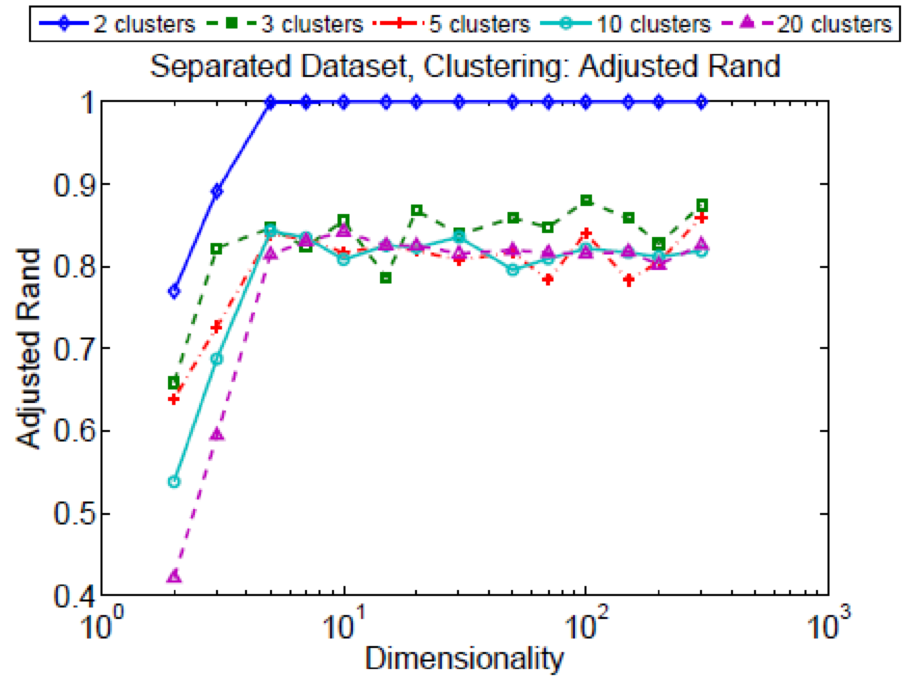


(b) Well-separated clusters

Sensitivity of the Average Quality Assessment: Adjusted Rand on *K*-Means



(a) Overlapping clusters



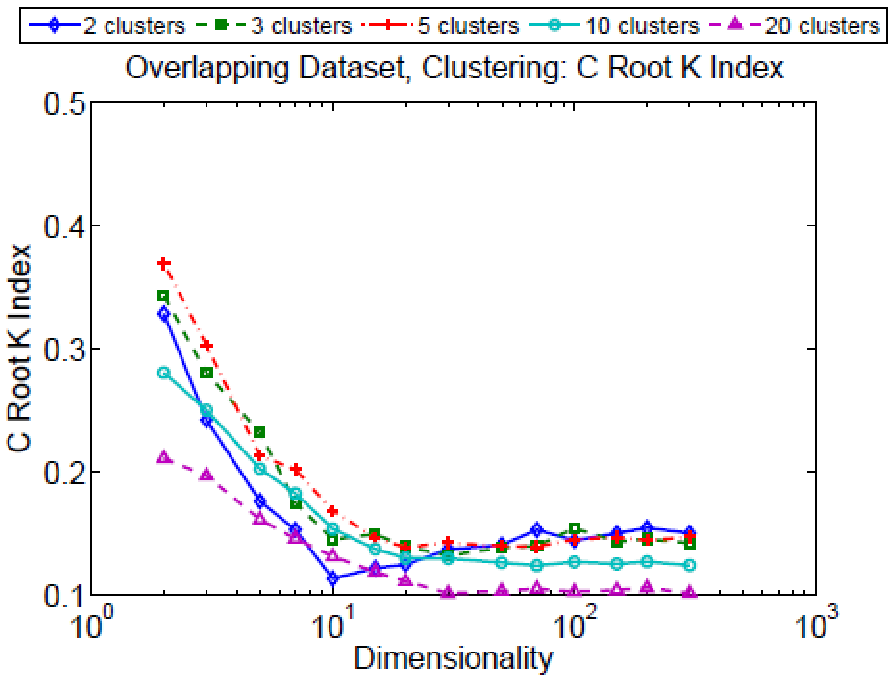
(b) Well-separated clusters

Sensitivity of the Average Quality Assessment

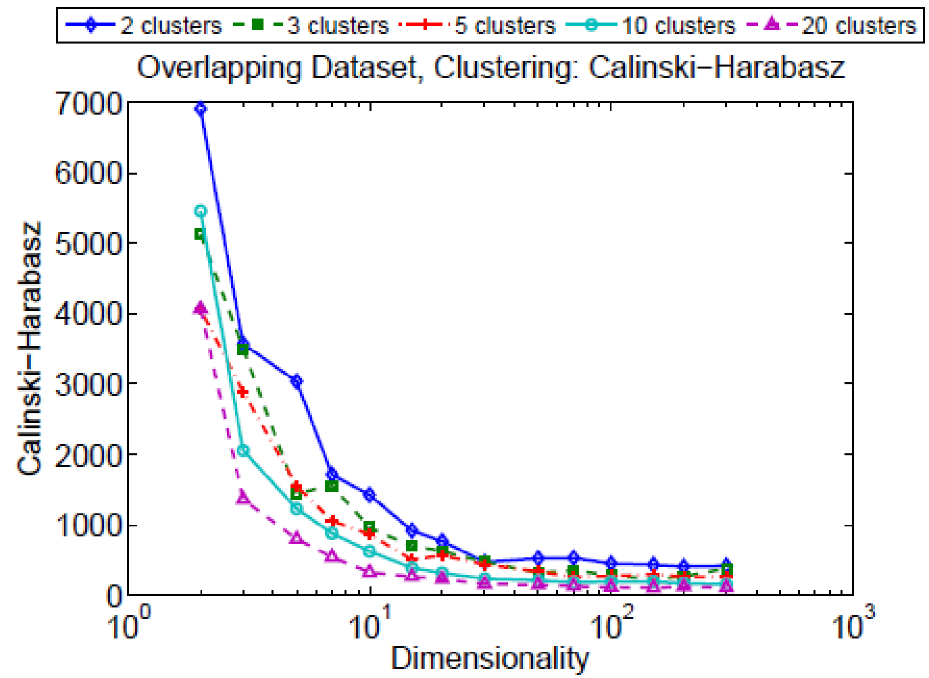
Evaluation of ***K*-means**

- However, the internal indexes behave in all sorts of ways, esp. in the overlapping cluster setting
- Some indexes robust w.r.t. ground truth, like G+ and Tau, still give consistent scores across dimensionalities
- Others that were robust, now give better scores to low-dimensional configurations ($C\sqrt{K}$, Calinski-Harabasz)
- Some indexes that increased with dimensionality on ground truth, now decrease (Silhouette)
- Point biserial and Hubert's statistic are U-shaped

Sensitivity of the Average Quality Assessment: $C\sqrt{K}$ and Calinski-Harabasz on *K*-Means

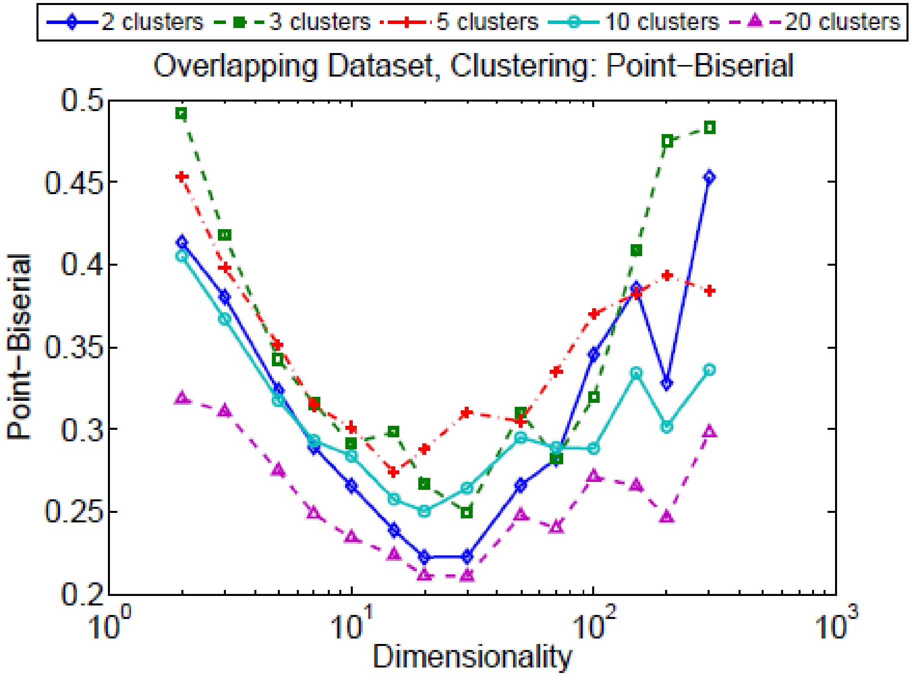


(a) $C\sqrt{K}$, Overlapping clusters

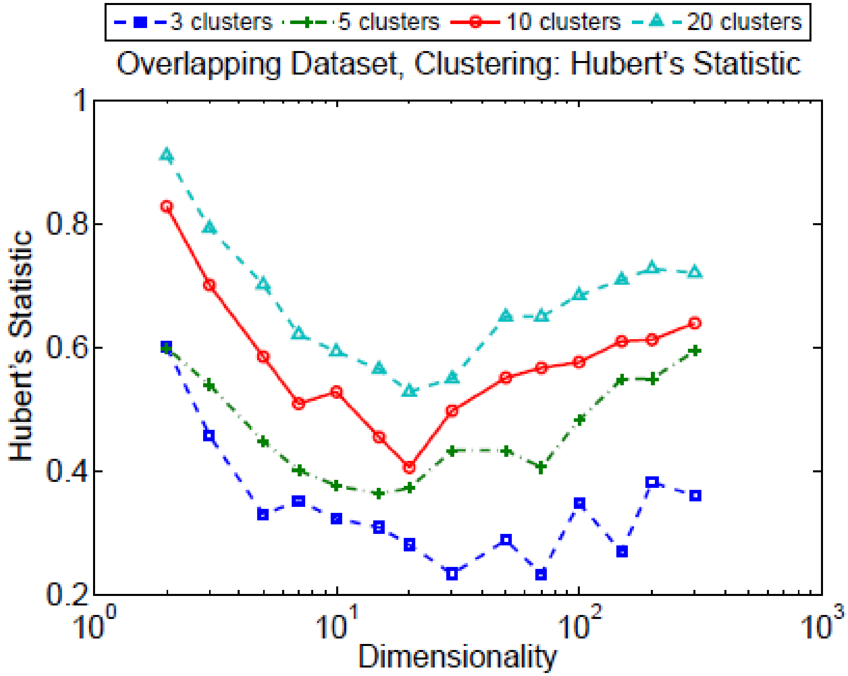


(b) Calinski-Harabasz, Overlapping clusters

Sensitivity of the Average Quality Assessment: Point-biserial and Hubert's Statistic on *K*-Means



(a) Point-biserial, Overlapping clusters

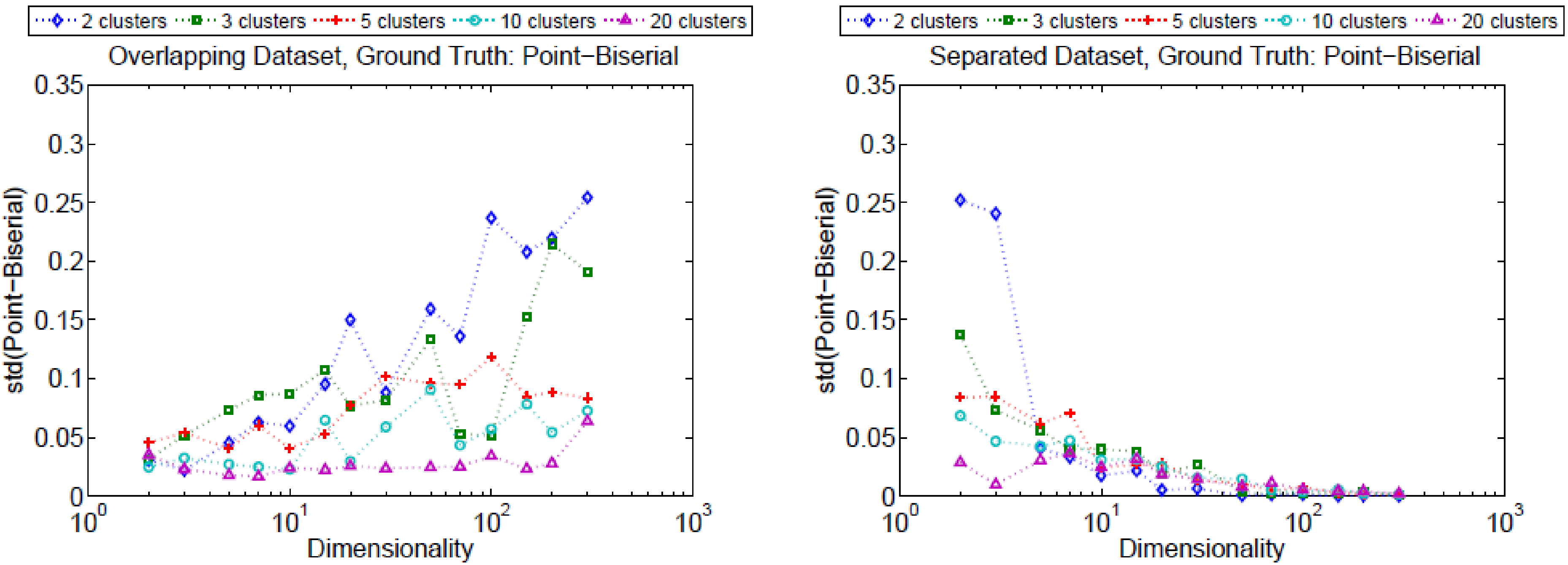


(b) Hubert's-Statistic, Overlapping clusters

Stability of Quality Assessment

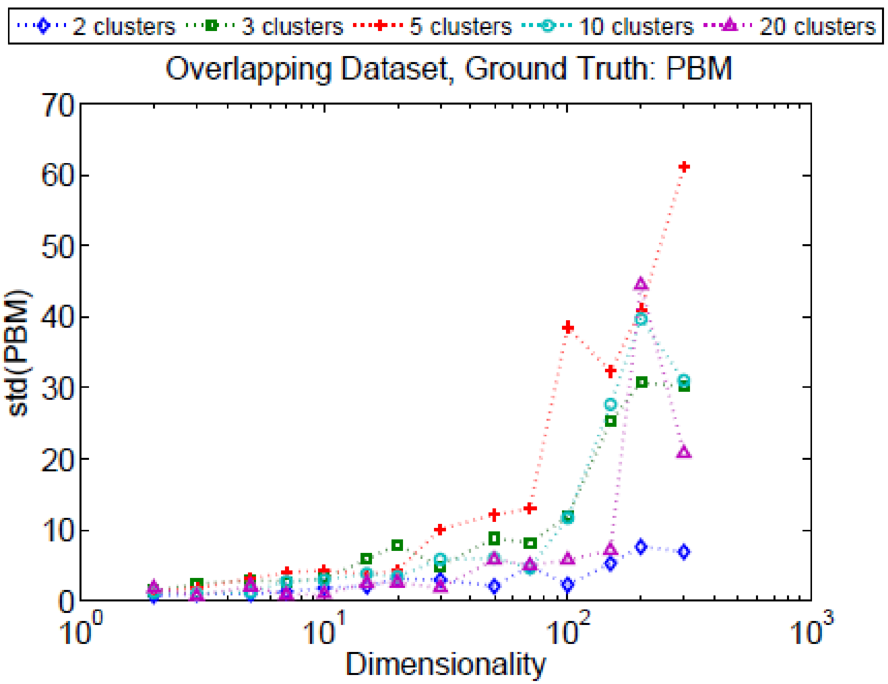
- Again, different indexes influenced in different ways in terms of score standard deviation
- Ground truth evaluation
- Point biserial: std increases in overlapping setting, decreases in separated setting
- PBM: std increases in both settings
- G_+, Tau, isolation index: std relatively stable

Stability of Quality Assessment: Point Biserial on Ground Truth

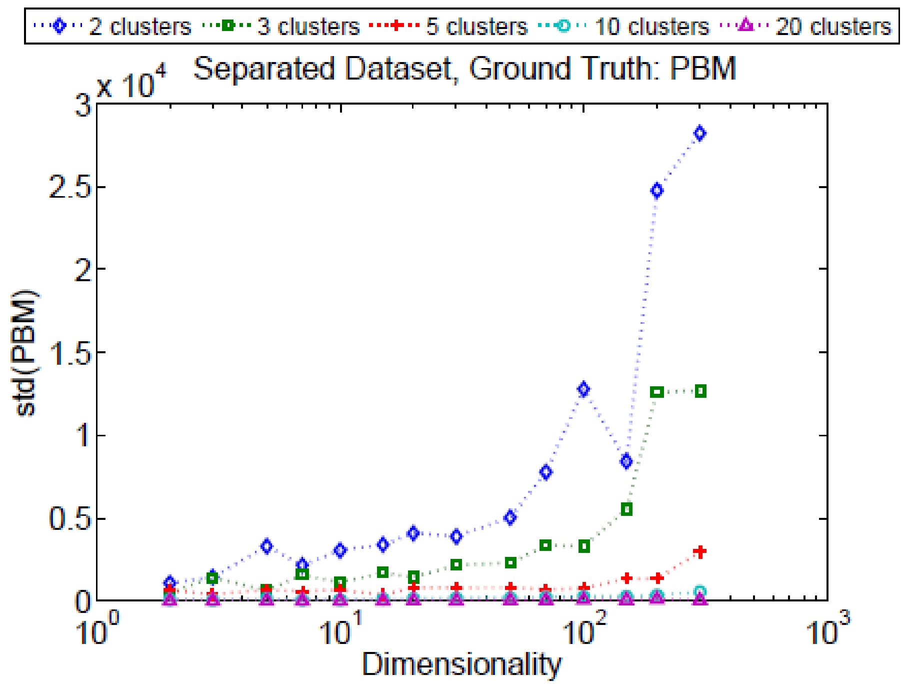


(a) Overlapping clusters (b) Well-separated clusters

Stability of Quality Assessment: PBM on Ground Truth

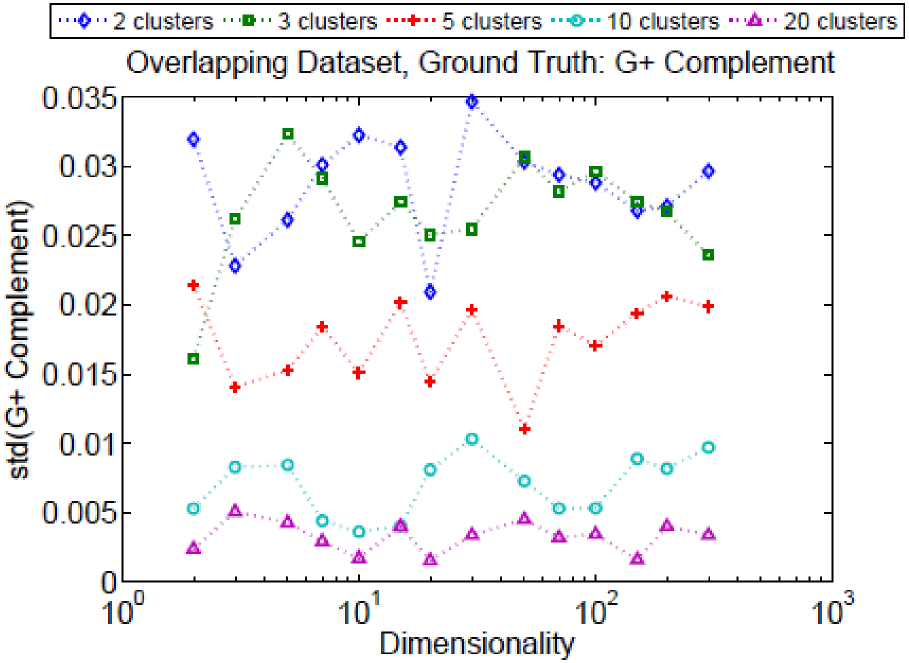


(a) Overlapping clusters

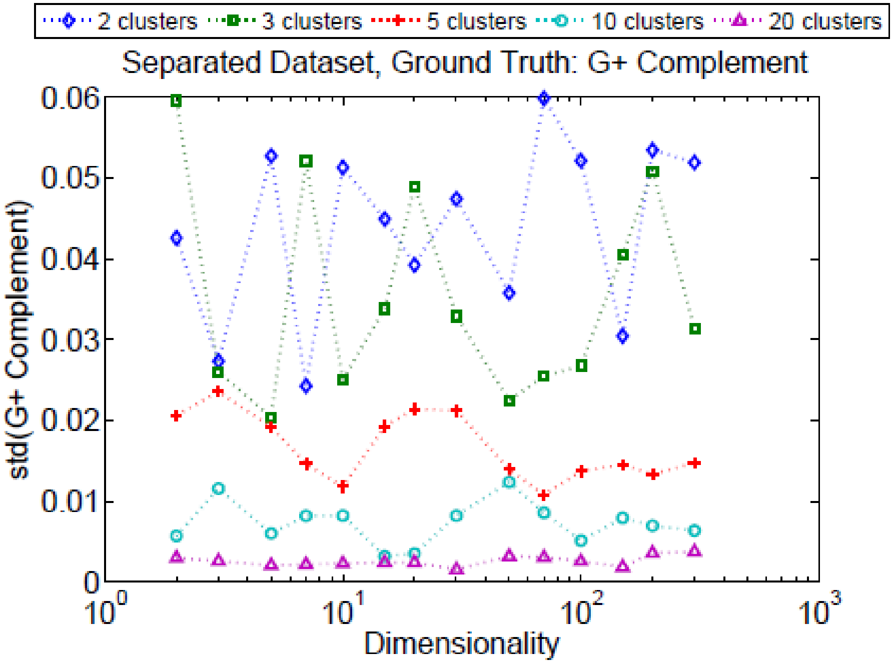


(b) Well-separated clusters

Stability of Quality Assessment: G_+ Complement on Ground Truth



(a) Overlapping clusters



(b) Well-separated clusters

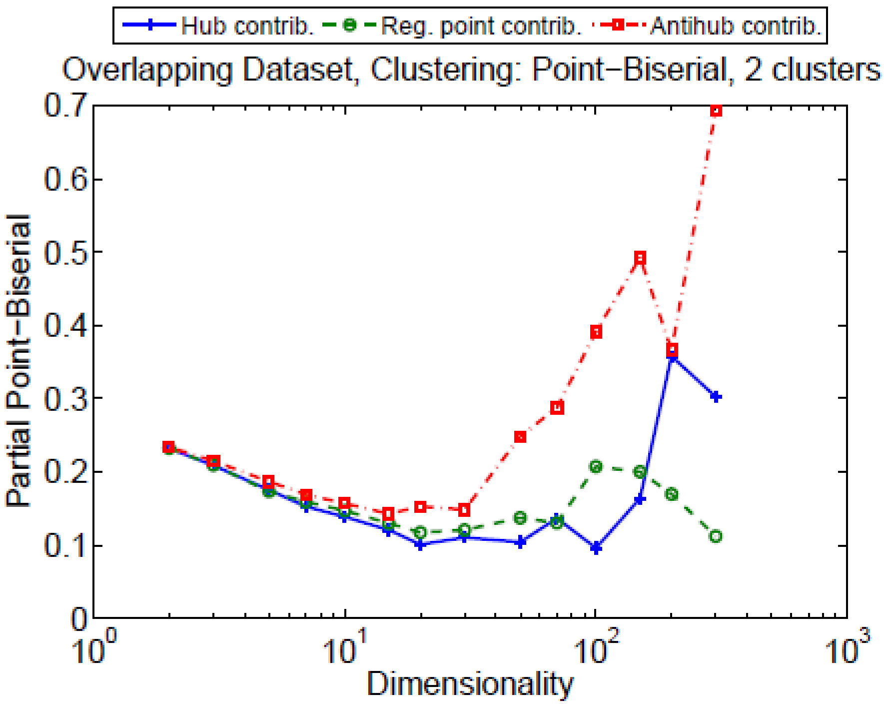
Outline

- Introduction
  - Curse of dimensionality, clustering quality indexes, distance concentration, hubness
- Clustering quality indexes: an overview
  - Internal indexes
  - External indexes
- Clustering evaluation in many dimensions
  - Experimental protocol
  - Sensitivity to increasing dimensionality
    - Sensitivity of the average quality assessment
    - Stability of quality assessment
  - Influence of hubs
- Conclusion and perspectives

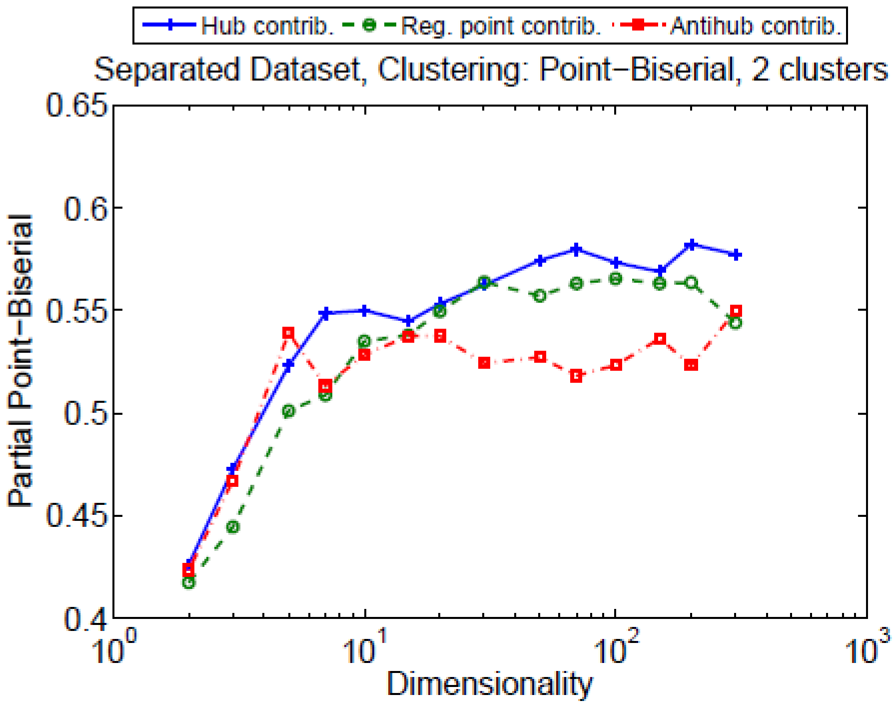
Influence of Hubs

- Hubs can cluster poorly by lowering between-cluster distance (esp. in cases when *K* is high)
- Demonstrated in our previous work for the Silhouette index [Radovanović et al. JMLR'10, Tomašev et al. TKDE'14]
- Here, we label points as hubs, regular points and anti-hubs by dividing the data set into three equal parts in the order of decreasing N_k score
- We express partial contributions of hubs, regular points, anti-hubs to various clustering indexes
- Whether hubs contribute substantially more or less than regular points for an index might affect the robustness of the index and its sensitivity to increasing dimensionality

Influence of Hubs

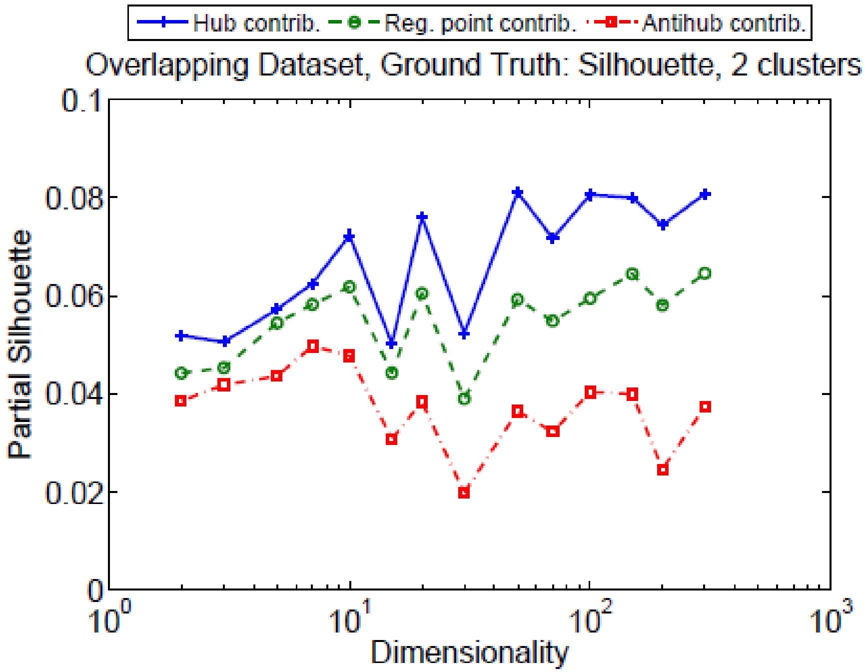


(a) Overlapping clusters

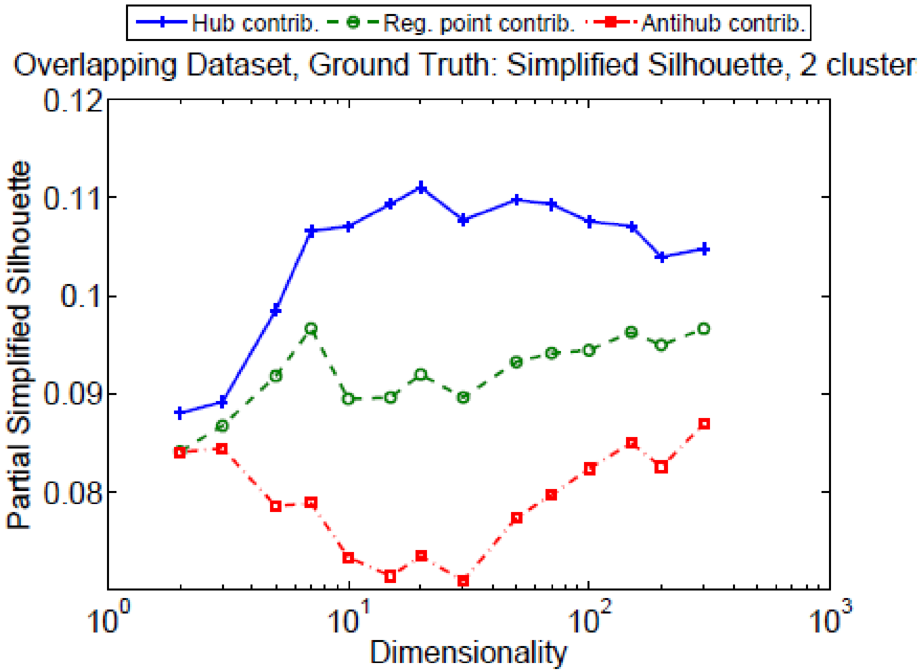


(b) Well-separated clusters

Influence of Hubs

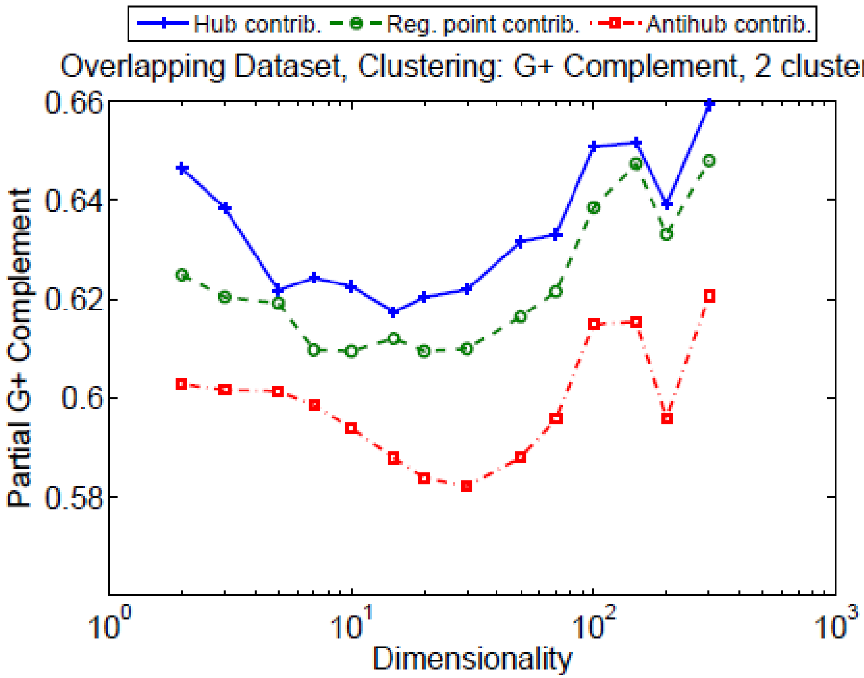


(a) Silhouette index

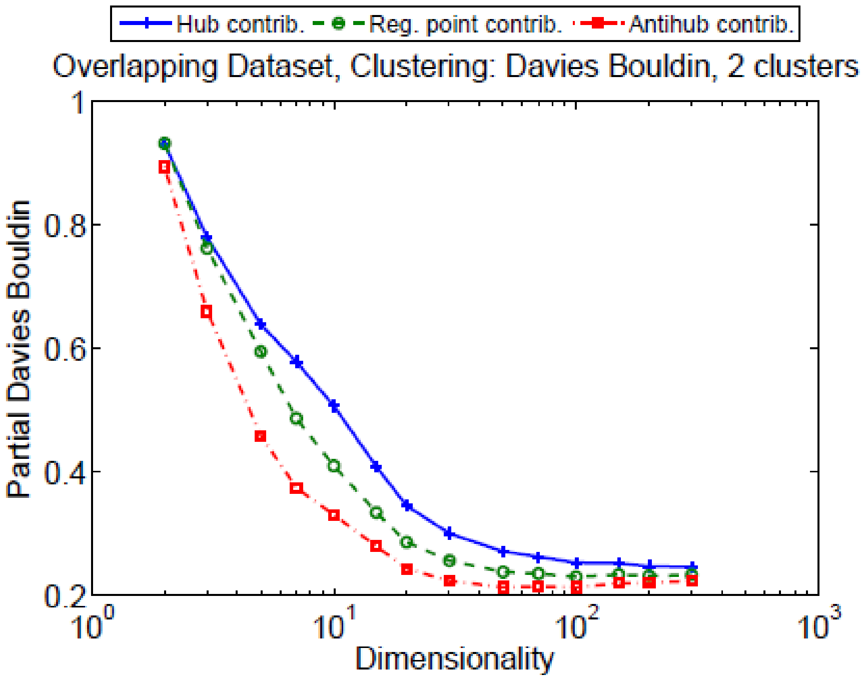


(b) Simplified Silhouette index

Influence of Hubs

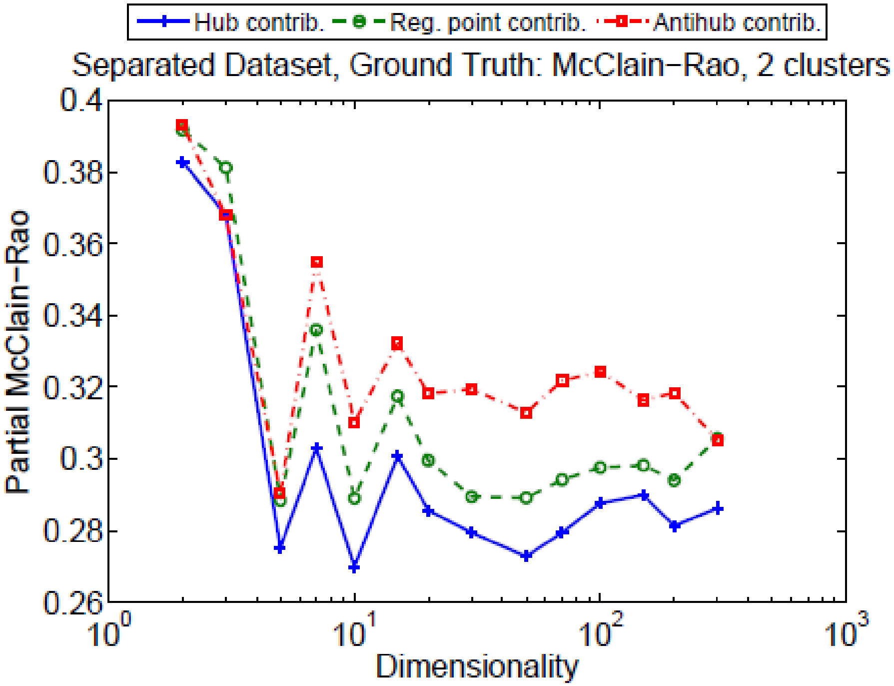


(a) $\bar{G}_+$ index, Overlapping clusters

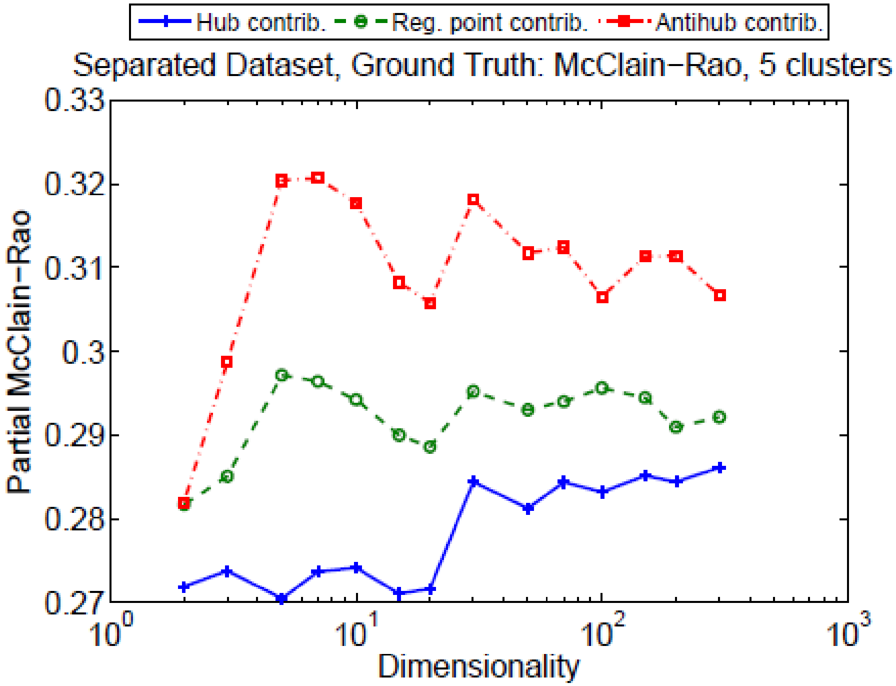


(b) Davies-Bouldin, Well-separated clusters

Influence of Hubs



(a) McClain-Rao index, 2 clusters



(b) McClain-Rao index, 5 clusters

Outline

- Introduction
  - Curse of dimensionality, clustering quality indexes, distance concentration, hubness
- Clustering quality indexes: an overview
  - Internal indexes
  - External indexes
- Clustering evaluation in many dimensions
  - Experimental protocol
  - Sensitivity to increasing dimensionality
    - Sensitivity of the average quality assessment
    - Stability of quality assessment
  - Influence of hubs
- Conclusion and perspectives

Conclusion and Perspectives

- Important to understand the behavior of clustering quality indexes in challenging contexts, like high dimensionality
- We showed that different indexes are influenced in different ways by increasing dimensionality
  - Average quality value (**bias**)
  - Stability of quality score (**variance**)
- What we have are initial results showing that selecting an appropriate index for high-dimensional data clustering is non-trivial and should be approached carefully
- For meaningful cross-index comparison, data dimensionality needs to be taken into account, otherwise results can simply be an artifact of dimensionality

Conclusion and Perspectives

- Hard to give general recommendations, but **G_+**, **Tau** and (to a lesser extent) **isolation index** showed best (in)sensitivity and stability across the board, w.r.t. dimensionality
- All indexes are sensitive to the number of clusters
- We used synthetic data, since it was easy to control the parameters
- A detailed study should be done on real data, by using repeated sub-sampling of larger high-dimensional datasets
  - Not many benchmark datasets with ground truth
- Better handling of hubs may result in better overall clustering quality: this could be incorporated into new/extended indexes

References

M. Radovanović et al. Nearest neighbors in high-dimensional data: The emergence and influence of hubs. In Proc. 26th Int. Conf. on Machine Learning (ICML), pages 865–872, 2009.

M. Radovanović et al. Hubs in space: Popular nearest neighbors in high-dimensional data. Journal of Machine Learning Research 11:2487–2531, 2010.

J.-J. Aucouturier and F. Pachet. A scale-free distribution of false positives for a large class of audio similarity measures. Pattern Recognition 41(1):272–284, 2007.

G. Doddington et al. SHEEP, GOATS, LAMBS and WOLVES: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. In Proc. 5th Int. Conf. on Spoken Language Processing (ICSLP), 1998. Paper 0608.

A. Hicklin et al. The myth of goats: How many people have fingerprints that are hard to match? Internal Report 7271, National Institute of Standards and Technology (NIST), USA, 2005.

H. Jegou et al. A contextual dissimilarity measure for accurate and efficient image search. In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1–8, 2007.

H. Jegou et al. Accurate image search using the contextual dissimilarity measure. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(1):2–11, 2010.

D. François et al. The concentration of fractional distances. IEEE Transactions on Knowledge and Data Engineering 19(7):873–886, 2007.

K. S. Beyer et al. When is “nearest neighbor” meaningful? In Proc. 7th Int. Conf. on Database Theory (ICDT), pages 217–235, 1999.

C. C. Aggarwal and P. S. Yu. Outlier detection for high dimensional data. In Proc. 27th ACM SIGMOD Int. Conf. on Management of Data, pages 37–46, 2001.

P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20:53–65, 1987.

L. Vendramin et al. Relative clustering validity criteria: A comparative overview. Statistical Analysis and Data Mining 3(4):209–235, 2010.

J. C. Dunn. Well-separated clusters and optimal fuzzy partitions. Journal of Cybernetics 4(1):95–104, 1974.

D. L. Davies and D. W. Bouldin. A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence 1(2):224–227, 1979.

E. J. Pauwels and G. Frederix. Cluster-based segmentation of natural scenes. In: Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV), vol. 2, pages 997–1002, 1999.

L. Hubert and J. Schultz. Quadratic assignment as a general data-analysis strategy. British Journal of Mathematical and Statistical Psychologie 29:190–241, 1976.

D. A. Ratkowsky and G. N. Lance. A criterion for determining the number of groups in a classification. Australian Computer Journal 10:115–117, 1978.

T. Calinski and J. Harabasz. A dendrite method for cluster analysis. Communications in Statistics, 3, no. 1:1–27, 1974.

L. Goodman and W. Kruskal. Measures of associations for cross-validations. Journal of the American Statistical Association 49:732–764, 1954.

F. B. Baker and L. J. Hubert. Measuring the power of hierarchical cluster analysis. Journal of the American Statistical Association 70:31–38, 1975.

F. J. Rohlf. Methods of comparing classifications. Annual Review of Ecology and Systematics 5:101–113, 1974.

M. Halkidi et al. On clustering validation techniques. Journal of Intelligent Information Systems 17(2-3):107–145, 2001.

J. O. McClain and V. R. Rao. Clustisz: A program to test for the quality of clustering of a set of objects. Journal of Marketing Research 12:456–460, 1975.

S. Bandyopadhyay et al. Validity index for crisp and fuzzy clusters. Pattern Recognition 37:487–501, 2004.

G. W. Milligan. A Monte Carlo study of thirty internal criterion measures for cluster analysis. Psychometrika 46(2):187–199, 1981.

S. C. Sharma. Applied Multivariate Techniques. John Wiley and Sons, 1996.

M. G. Kendall and J. D. Gibbons, Rank Correlation Methods, London, UK, Edward Arnold, 1990.

R. J. G. B. Campello and E. R. Hruschka, On comparing two sequences of numbers and its applications to clustering analysis. Information Sciences 179(8):1025–1039, 2009.

W. M. Rand. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66(336):846–850, 1971.

L. Hubert and P. Arabie. Comparing partitions. Journal of Classification 2(1):193–218, 1985.

E. B. Fowlkes and C. L. Mallows. A method for comparing two hierarchical clusterings. Journal of the American Statistical Association 78(383):553–569, 1983.

N. Tomašev et al. The role of hubness in clustering high-dimensional data. IEEE Transactions on Knowledge and Data Engineering 26(3):739–751, 2014.