Spectral Clustering

Kainat Khowaja

Wolfgang Karl Härdle

IRTG 1792
High Dimensional Non-Stationary Time Series
Humboldt-Universität zu Berlin
IRTG1792.HU-Berlin.de

Motivation

- Computer Vision
  - Image segmentation

Original

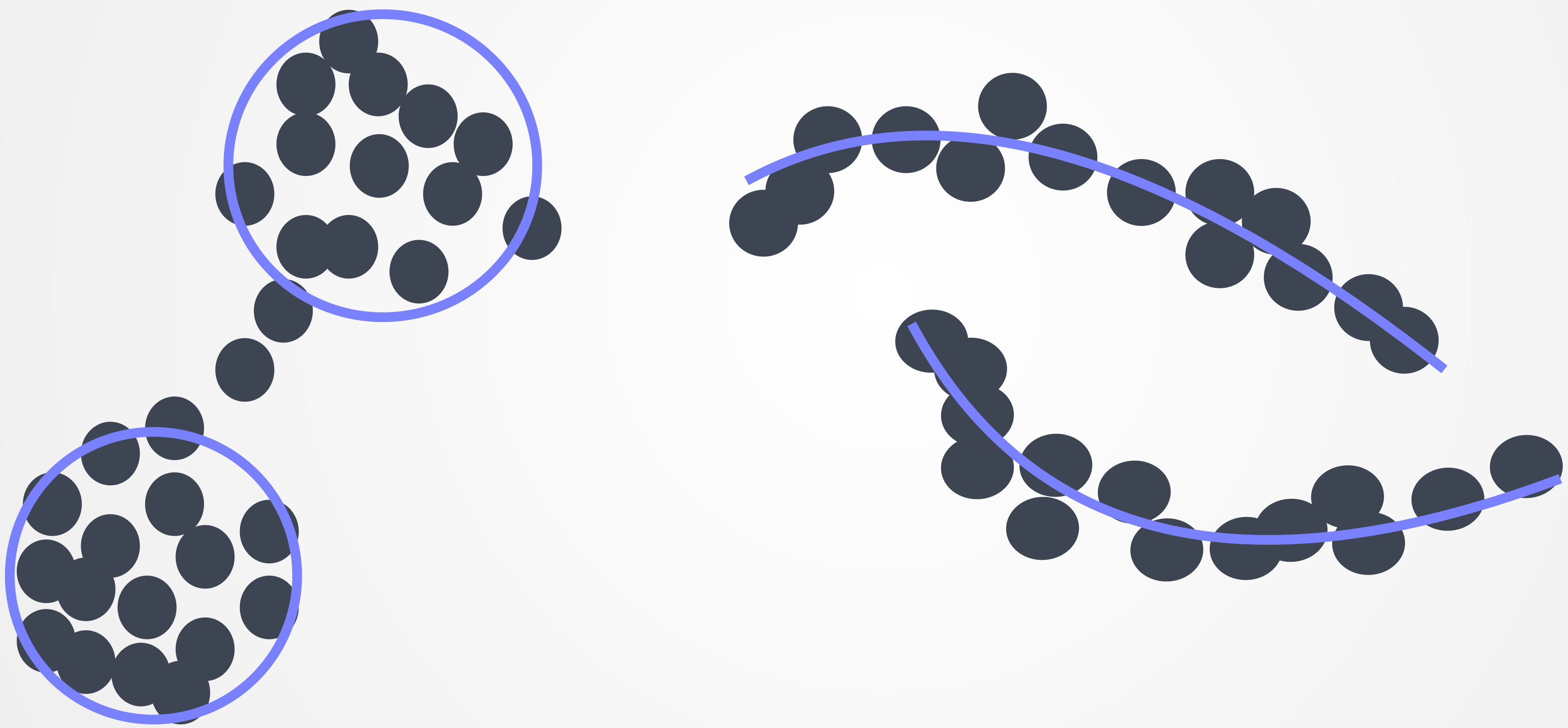
2 Clusters

4 Clusters

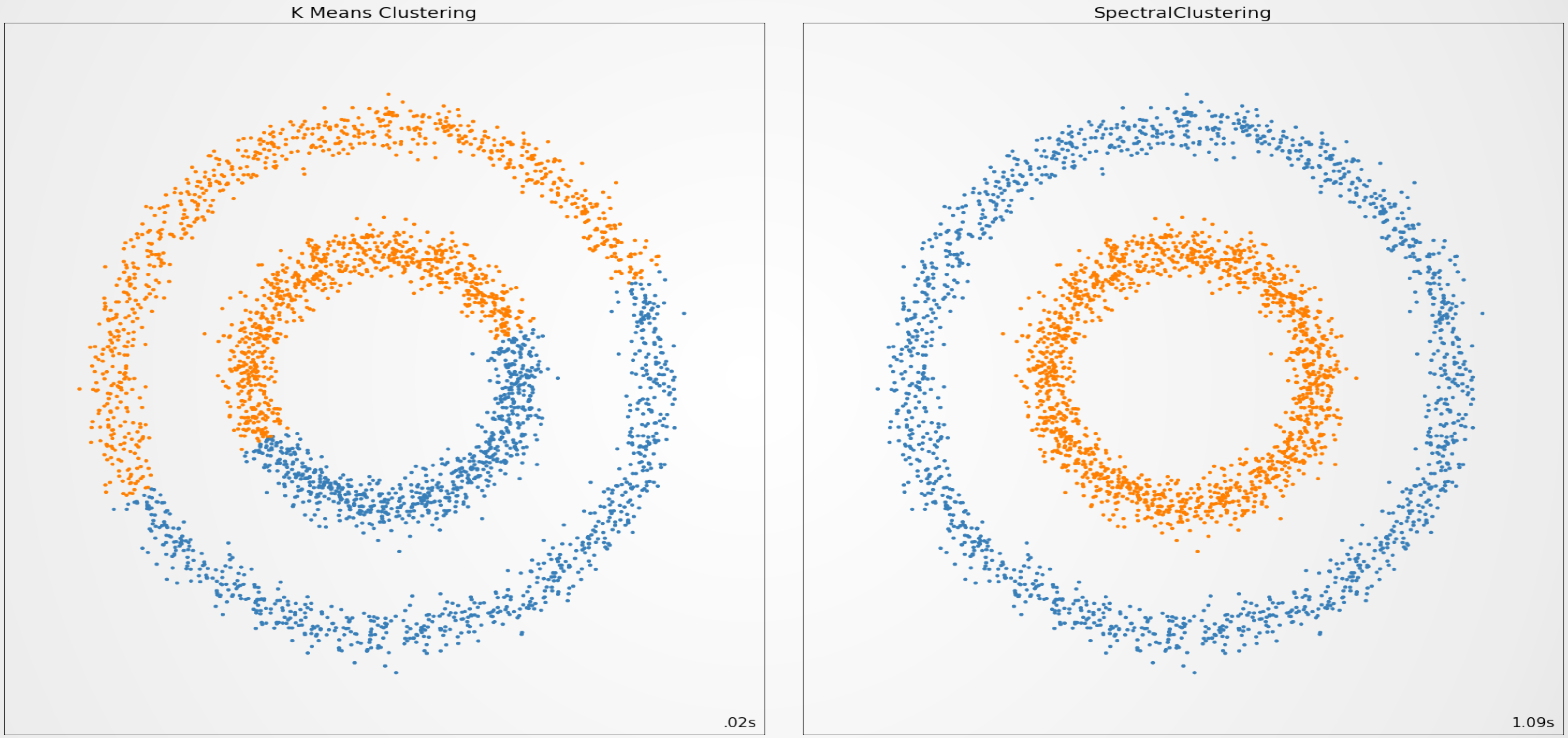
8 Clusters

Idea

- Clustering techniques are usually based on either
  - Compactness, e.g. *k*-means
  - Connectivity, e.g. spectral clustering



Why Spectral Clustering?

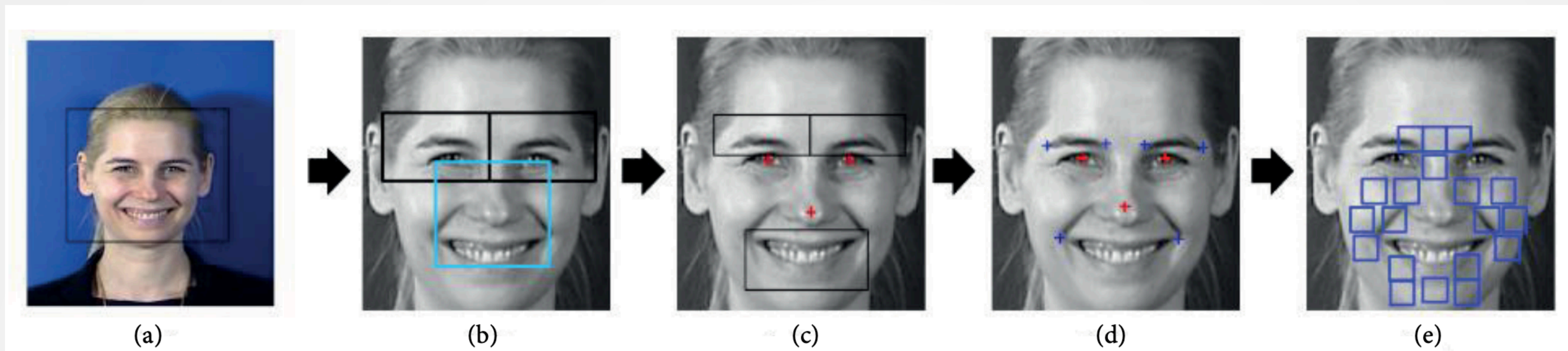


Clustering_algorithm_comparison

Poor performance of neural networks for similar classification problem
(motivated by BS)

Applications

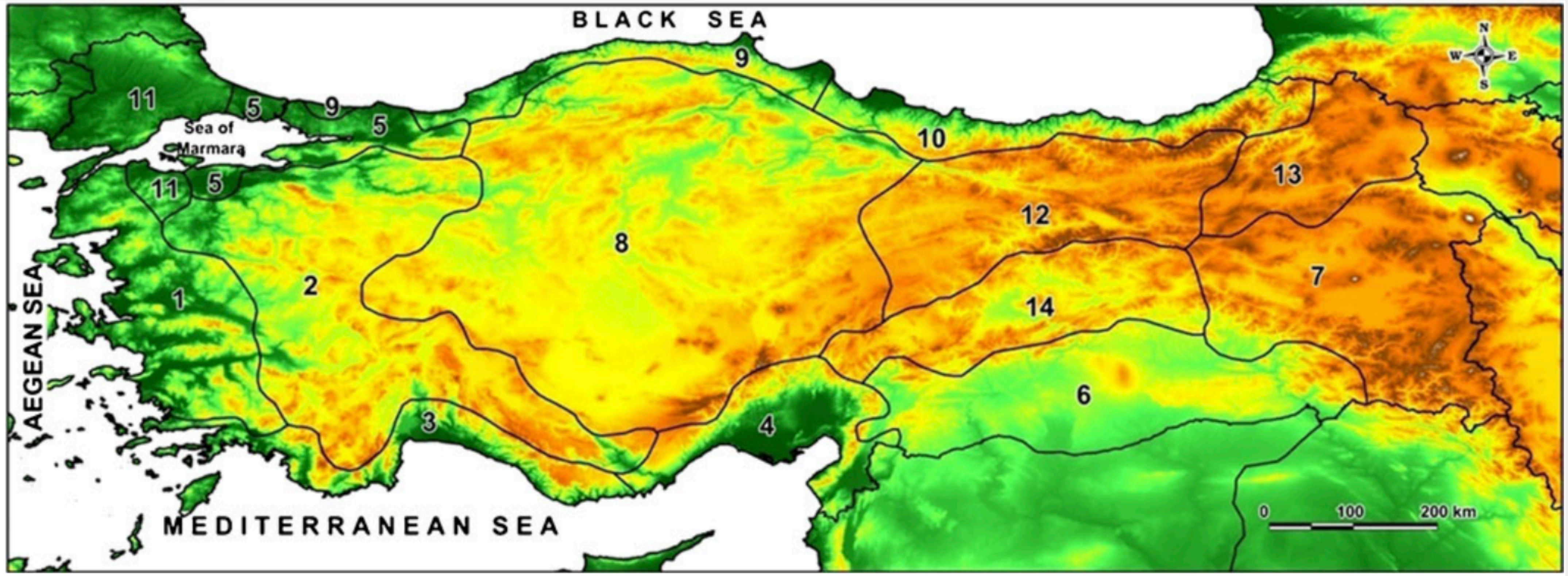
- Facial expression recognition (Cai X. et al 2011, Bian J. et al 2019)



- Image segmentation in medicine (detection of tumors (examples in Koch Institute Gallery), abnormalities in ultrasound images (Mohammad S & Salama M 2007)

Self-Assembly in a Next Generation Polymer

- Clustering of climatic regions in Turkey (Iyigun C. et al 2013)

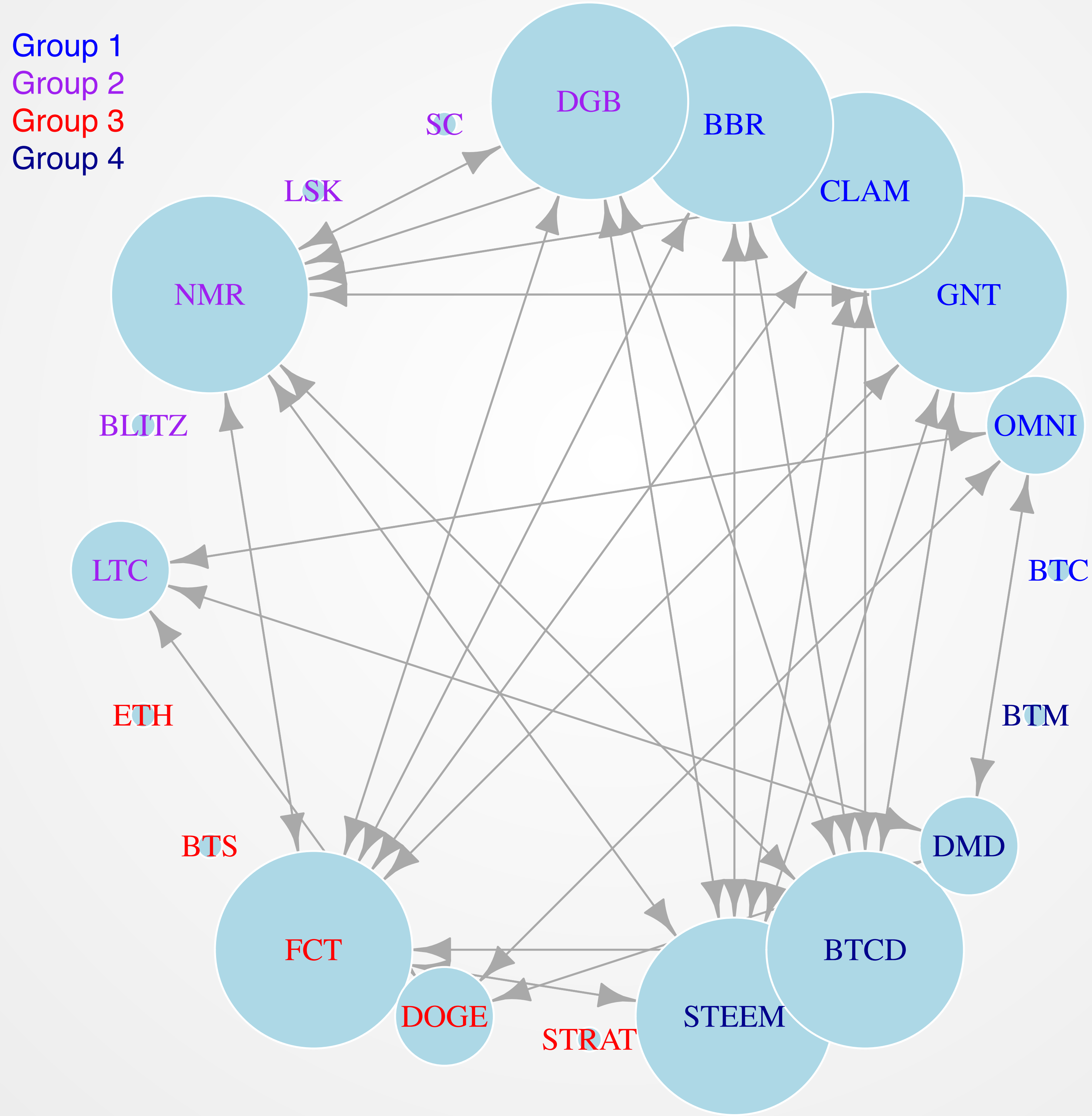


- More applications

  https://arxiv.org/pdf/1505.00477.pdf

Crypto Clustering

- Which Altcoins belong to which cluster? (Li G. et al 2019)

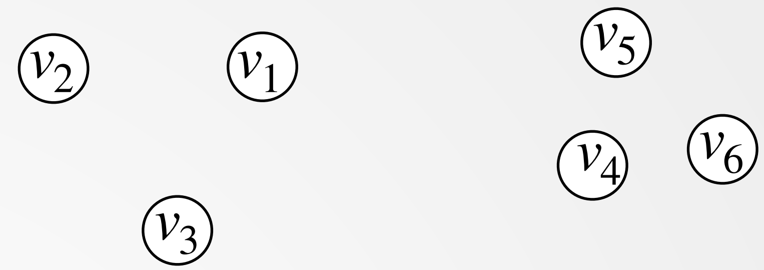


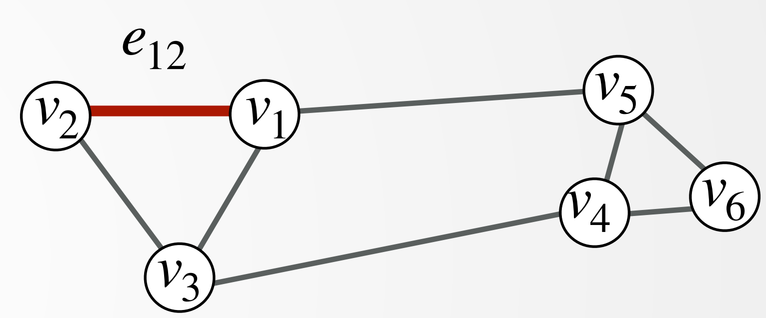
Outline

1. Motivation ✓
2. Graph Clustering
3. Spectral Clustering Algorithm
4. Example
5. References

Graph Clustering

- Data as graphs G=(V,E)
  - Vertex set $V = \{v_1, \ldots, v_n\}$
  - Edges represent similarity (ϵ neighbourhood, *k*-NNs, fully connected graph)



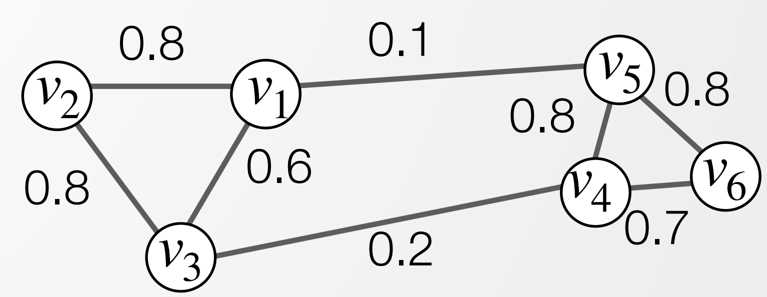


Similarity Graphs: local neighborhood relations between data points

E.g. Gaussian kernel similarity function

$$w_{ij} = e^{\frac{||v_i - v_j||^2}{2\sigma^2}}$$

size of nbd

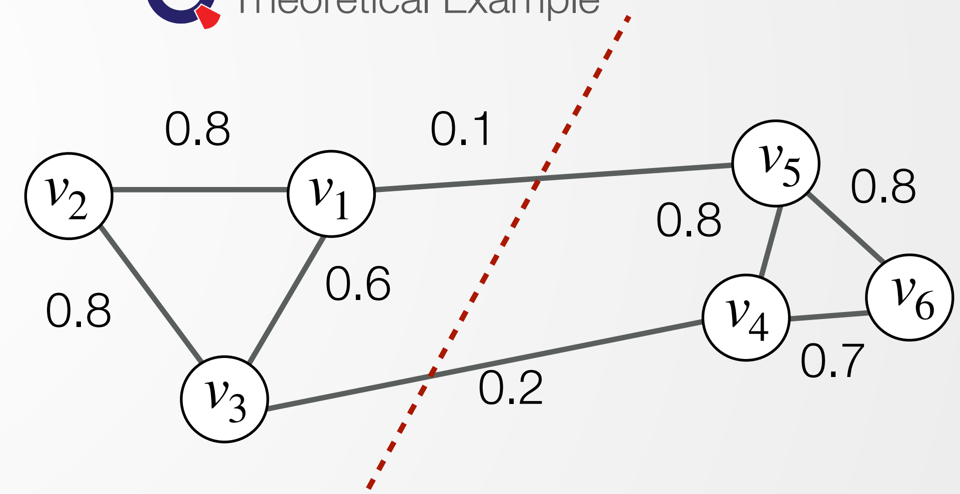


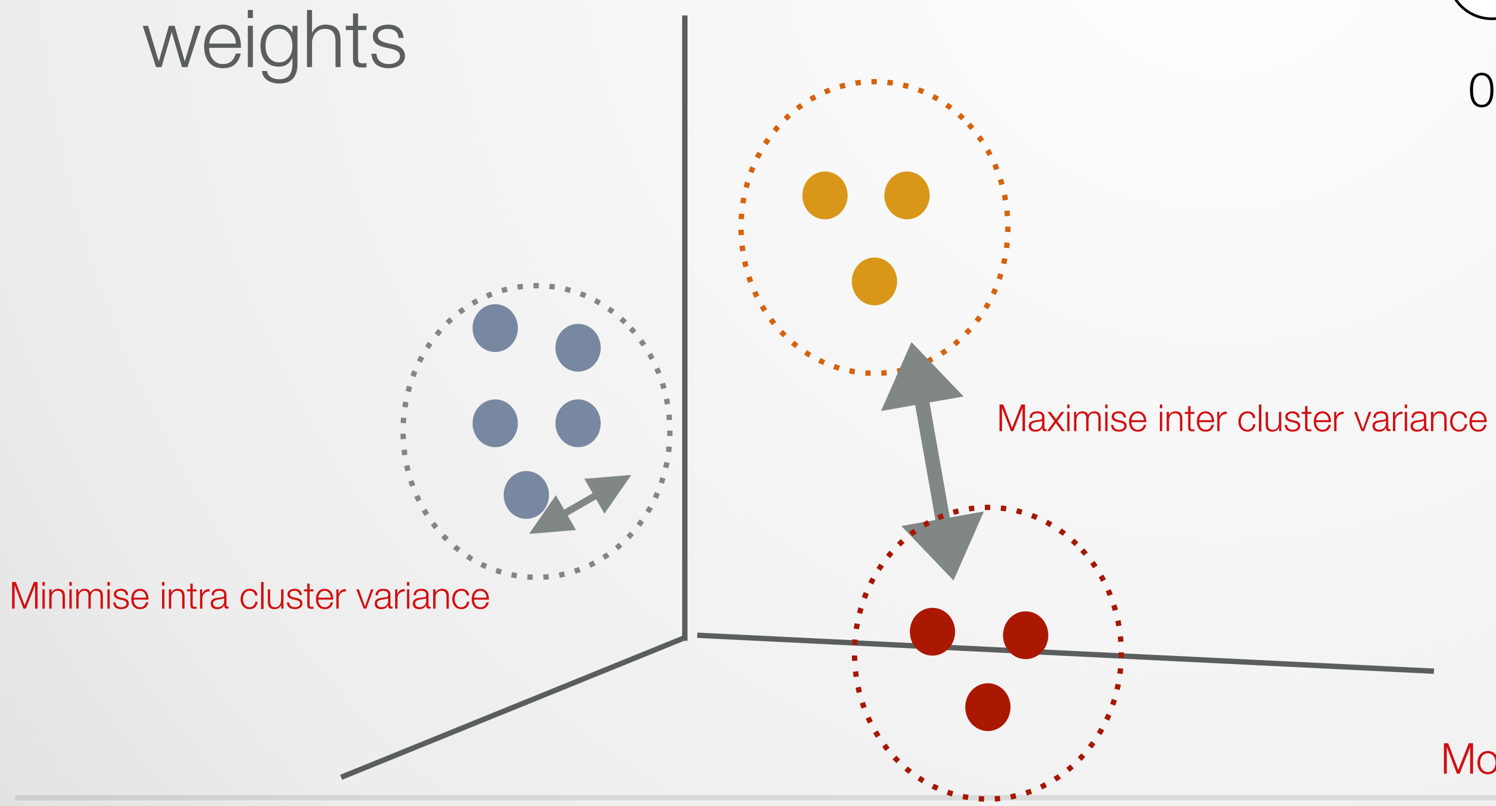
Graph Clustering

- Weighted adjacency matrix/ Similarity matrix
  - $A = (w_{ij}), i, j = 1, \ldots, n$ and $w_{ij} \geq 0$
- Partitions such as edges within clusters have higher weights and edges across sets have smaller weights

$$A = \begin{bmatrix} 0.0 & 0.8 & 0.6 & 0.0 & 0.1 & 0.0 \\ 0.8 & 0.0 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.6 & 0.8 & 0.0 & 0.2 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.0 & 0.8 & 0.7 \\ 0.1 & 0.0 & 0.0 & 0.8 & 0.0 & 0.8 \\ 0.0 & 0.0 & 0.0 & 0.7 & 0.8 & 0.0 \end{bmatrix}$$

Theoretical Example



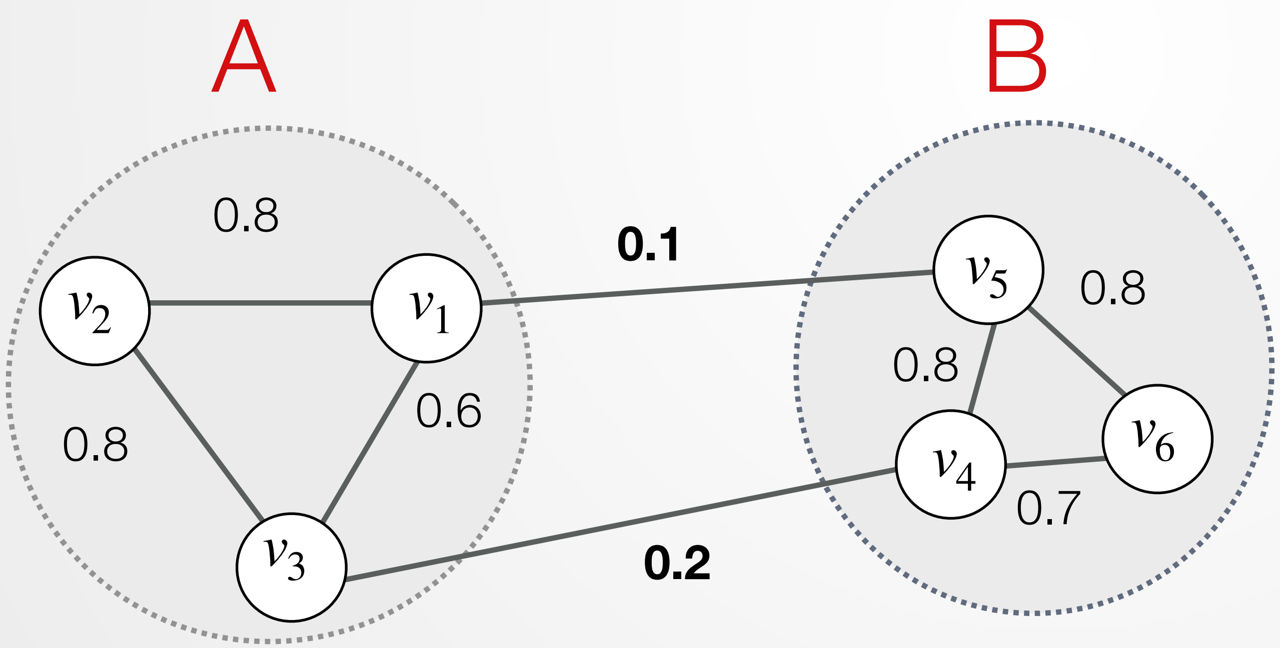


Motivated by EZ

Partitioning a graph into clusters

- Cut: Partition in disjoint $A, B \subset V$ such that weight of edges connecting vertices in A to vertices in B is minimum

$$cut(A,B) \stackrel{def}{=} \sum_{i \in A, j \in B} w_{ij}$$

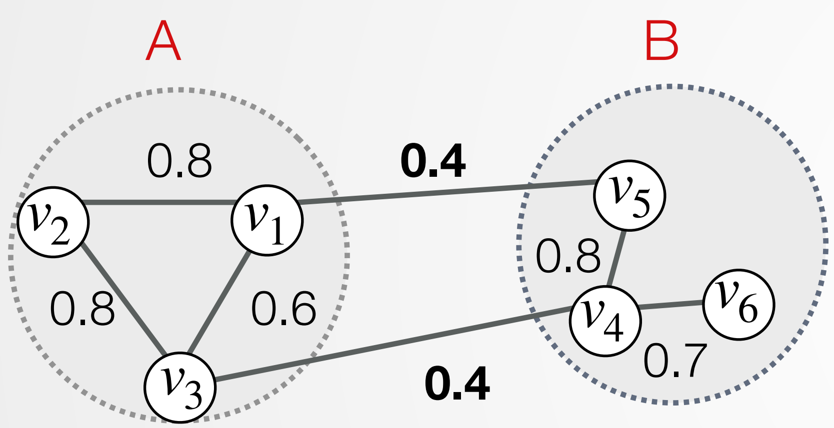


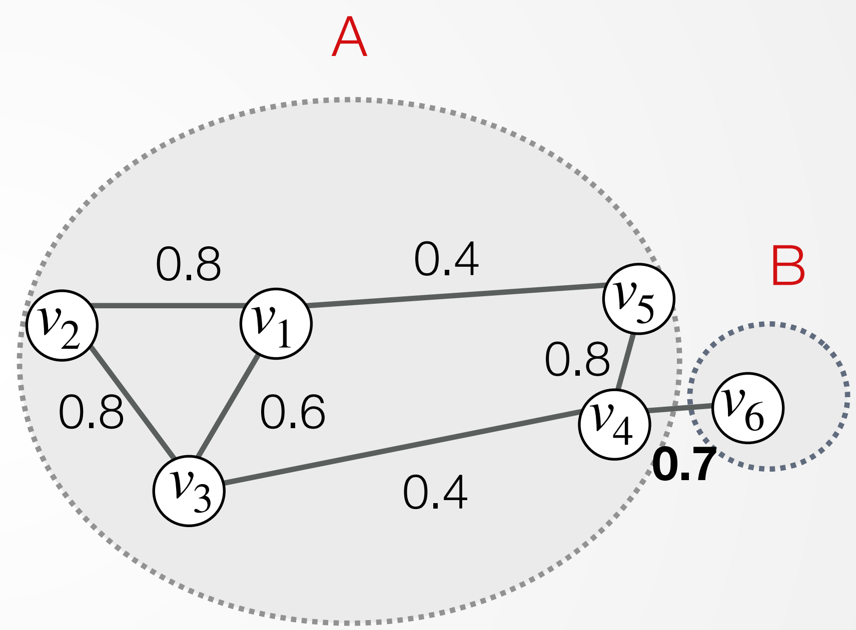
$$cut(A,B) = 0.1 + 0.2 = 0.3$$

Partitioning a graph into clusters

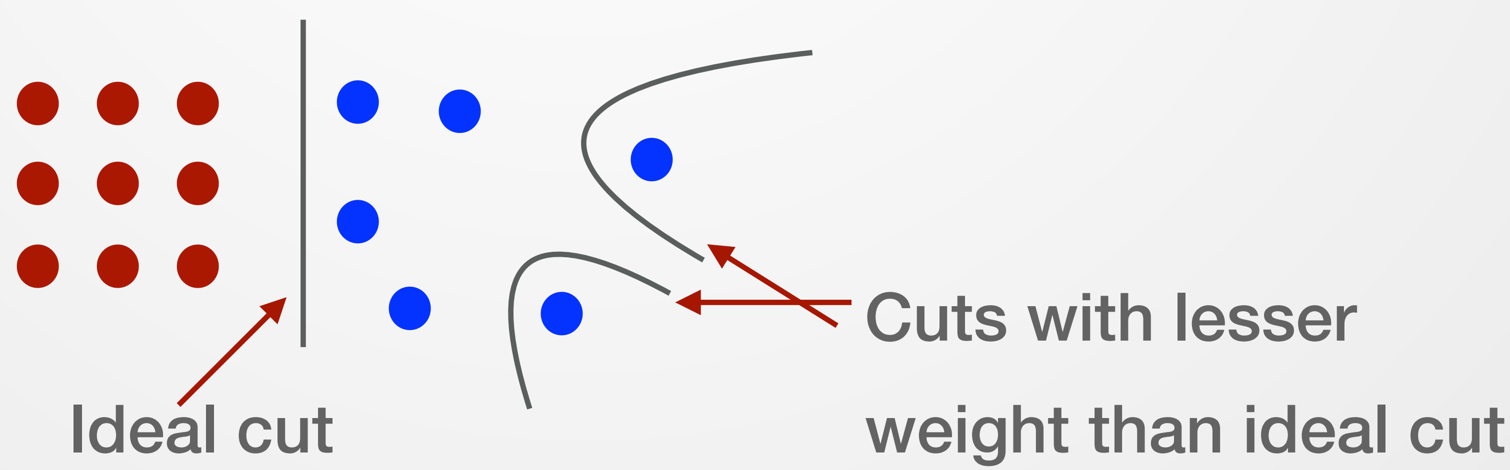
- Min cut: Given similarity graph with adjacency matrix A to be partitioned, minimize

$$cut(A_1, \ldots, A_k) \stackrel{def}{=} \sum_{i=1}^{k} cut(A_i, \bar{A}_i)$$





- Problem: separates individual vertex from the rest



Partitioning a graph into clusters

- Ratio cut : size measured by number of vertices $|A_i|$

$$RatioCut(A_1, \ldots, A_k) \stackrel{def}{=} \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{|A_i|}$$

- Normal cut: the size is measured by weights of its edges $vol(A_i)$

$$NormalCut(A_1, \ldots, A_k) \stackrel{def}{=} \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{vol(A_i)}$$

But NP-hard to solve!!

SC is a relaxation of these:

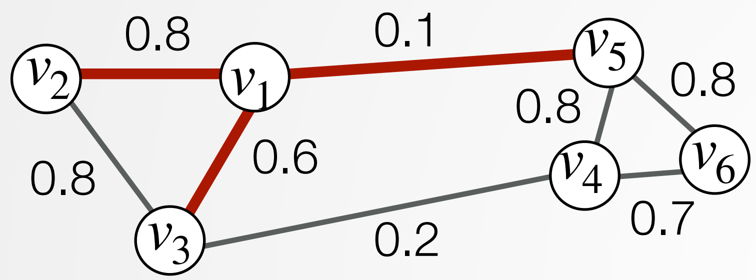
Relaxed Ratio cut → Unnormalized SC

Relaxed Normal cut → Normalized SC

Graph Laplacian

- Degree Matrix

$$D = diag(d_i = \sum_{j=1}^{n} w_{ij})$$



$$D = \begin{bmatrix} 1.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.6 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.6 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.7 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.7 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.5 \end{bmatrix}$$

- Unnormalized graph Laplacian

$$L_U = D - A$$

$$L_U = \begin{bmatrix} 1.5 & -0.8 & -0.6 & 0.0 & -0.1 & 0.0 \\ -0.8 & 1.6 & -0.8 & 0.0 & 0.0 & 0.0 \\ -0.6 & -0.8 & 1.6 & -0.2 & 0.0 & 0.0 \\ 0.0 & 0.0 & -0.2 & 1.7 & -0.8 & -0.7 \\ -0.1 & 0.0 & 0.0 & -0.8 & 1.7 & -0.8 \\ 0.0 & 0.0 & 0.0 & -0.7 & -0.8 & 1.5 \end{bmatrix}$$

- Normalized Graph Laplacian

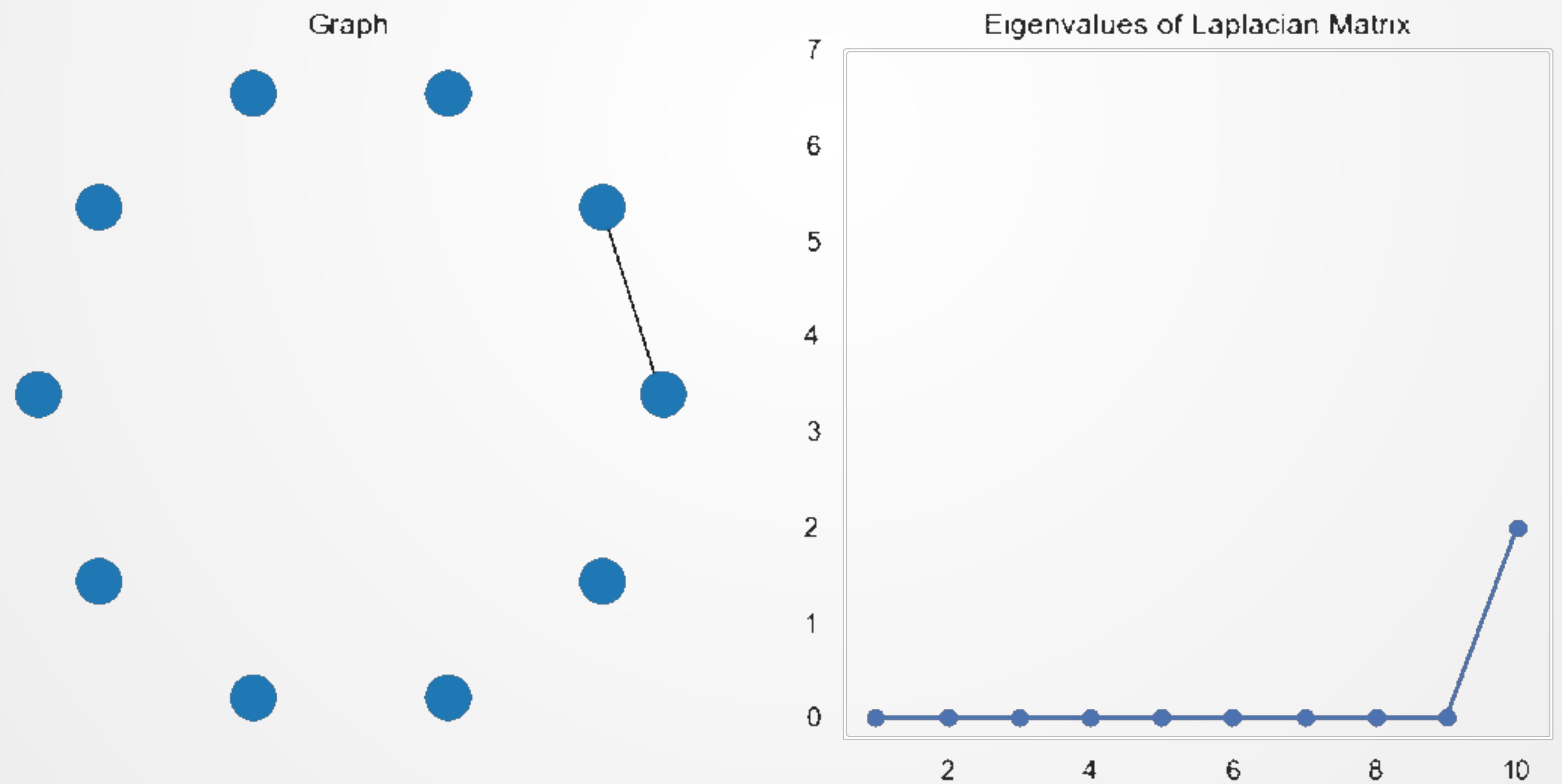
$$L_N = D^{-0.5} L_U D^{-0.5}$$

$$L_N = \begin{bmatrix} -0.41 & -0.41 & 0.65 & 0.31 & -0.38 & -0.12 \\ -0.41 & -0.44 & -0.01 & -0.31 & 0.71 & -0.22 \\ -0.41 & -0.37 & -0.64 & -0.05 & -0.39 & 0.37 \\ -0.41 & 0.37 & -0.34 & 0.46 & -0.00 & -0.61 \\ -0.41 & 0.41 & 0.17 & 0.31 & 0.35 & 0.65 \\ -0.41 & 0.45 & 0.18 & -0.72 & -0.29 & -0.09 \end{bmatrix}$$

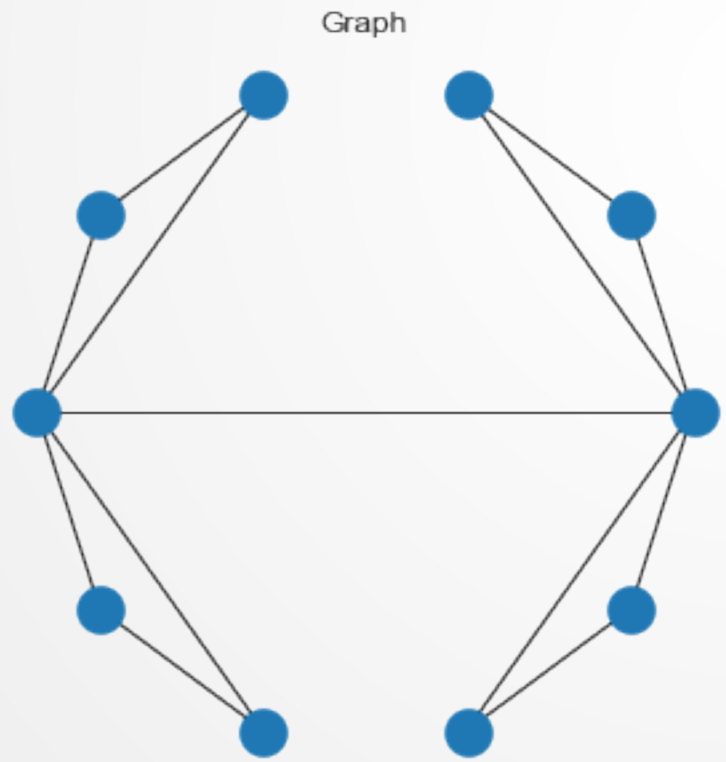
Theoretical Example

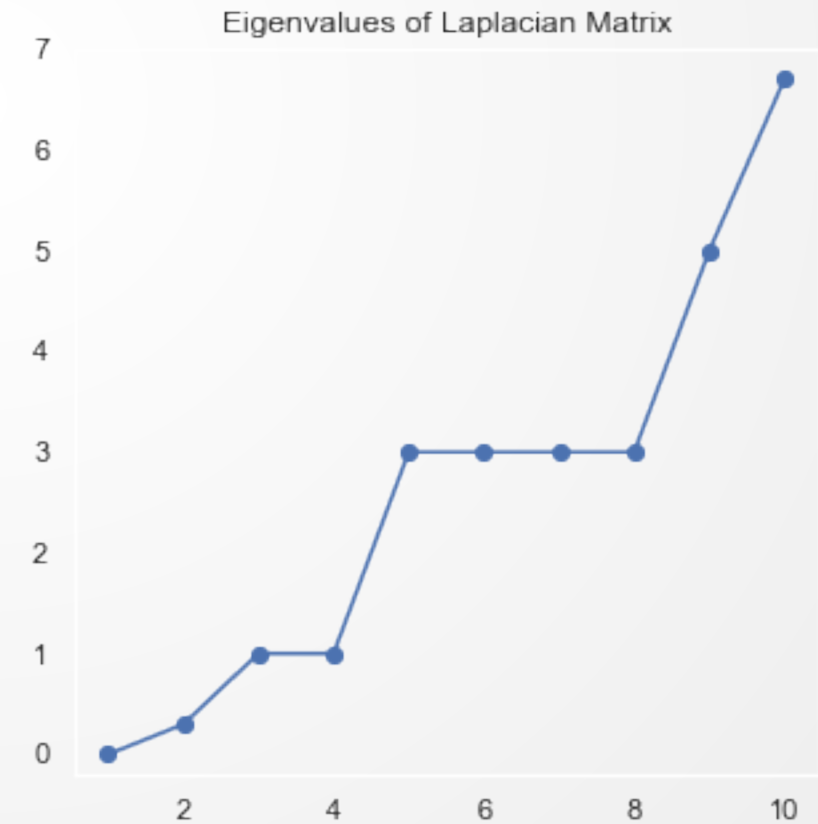
Graph Laplacian

- Laplacian has some beautiful properties
  - Symmetric and positive semi-definite
  - The smallest EVal of L is 0, EVec is constant one vector
  - Non-negative, real valued eigenvalues



- # zero eigenvalues = # connected components
- First non zero EVal “spectral gap”
- In previous example, if graph was densely connected, spectral gap would have been 10
- Smaller spectral gap: neater clusters (because close to disconnected components)





Graph Laplacian and Cuts

- For every vector $f \in \mathbb{R}^n$ we have

$$f^{\mathsf{T}} L_U f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} \left(f_i - f_j \right)^2$$

$$f^{\mathsf{T}} L_N f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} \left(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_i}} \right)^2$$

- Relationship with Ratio Cut and Normal Cut
  - Given $A \subset V$, define $f = (f_1, \ldots, f_n)^{\mathsf{T}} \in \mathbb{R}^n$ as

$$f_i = \begin{cases} \sqrt{|\bar{A}|/|A|} & if\ v_i \in A \\ -\sqrt{|A|/|\bar{A}|} & if\ v_i \in \bar{A} \end{cases}$$

Graph Laplacian and Cuts

$$f^{\mathsf{T}} L_U f = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}\left(f_i - f_j\right)^2$$

$$= \sum_{i\in A, j\in \bar{A}} w_{ij}\left(\sqrt{\frac{|\bar{A}|}{|A|}} + \sqrt{\frac{|A|}{|\bar{A}|}}\right)^2 + \sum_{i\in \bar{A}, j\in A} w_{ij}\left(-\sqrt{\frac{|\bar{A}|}{|A|}} - \sqrt{\frac{|A|}{|\bar{A}|}}\right)^2$$

$$= 2\ cut(A,\bar{A})\left(\frac{|\bar{A}|}{|A|} + \frac{|A|}{|\bar{A}|} + 2\right)$$

$$= 2\ cut(A,\bar{A})\left(\frac{|\bar{A}| + |A|}{|A|} + \frac{|A| + |\bar{A}|}{|\bar{A}|}\right)$$

$$= 2\,|V|\,.\,RatioCut(A,\bar{A})$$

Note: $\mathbf{1}\,.\,\sum_{i=1}^{n} f_i = \sum_{i\in A}\sqrt{\frac{|\bar{A}|}{|A|}} - \sum_{i\in \bar{A}}\sqrt{\frac{|A|}{|\bar{A}|}} = |A|\sqrt{\frac{|\bar{A}|}{|A|}} - |\bar{A}|\sqrt{\frac{|A|}{|\bar{A}|}} = 0$

and $\quad \|f\|^2 = \sum_{i=1}^{n} f_i^2 = A|\frac{|\bar{A}|}{|A|} + |\bar{A}|\frac{|A|}{|\bar{A}|} = |\bar{A}| + |A| = n$

Graph Laplacian and Cuts

- Problem:
  - $\min\limits_{A \subset V} f^{\top} L f$ subject to $f \perp \mathbf{1}$, f_i as defined, $\|f\| = \sqrt{n}$
- Solution: NP hard- discrete values of f_i
- Relaxation: Allow $f_i \in \mathbb{R}$
  - $\min\limits_{f \in \mathbb{R}^n} f^{\top} L f$ subject to $f \perp \mathbf{1}$, f_i as defined, $\|f\| = \sqrt{n}$
- Rayleigh-Ritz Theorem: solution given by Evec of second smallest EVal of L
- Analogous proceedings for $k \neq 2$ and normalized SC.

Spectral Bi-Partitioning Algorithm

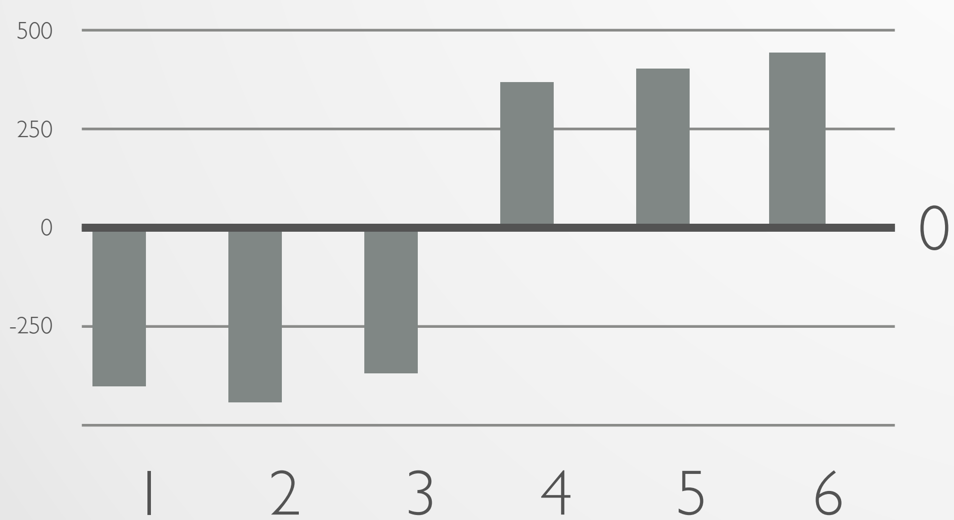
- $EVals = [0.000 \quad 0.188 \quad 2.084 \quad 2.285 \quad 2.469 \quad 2.573]$

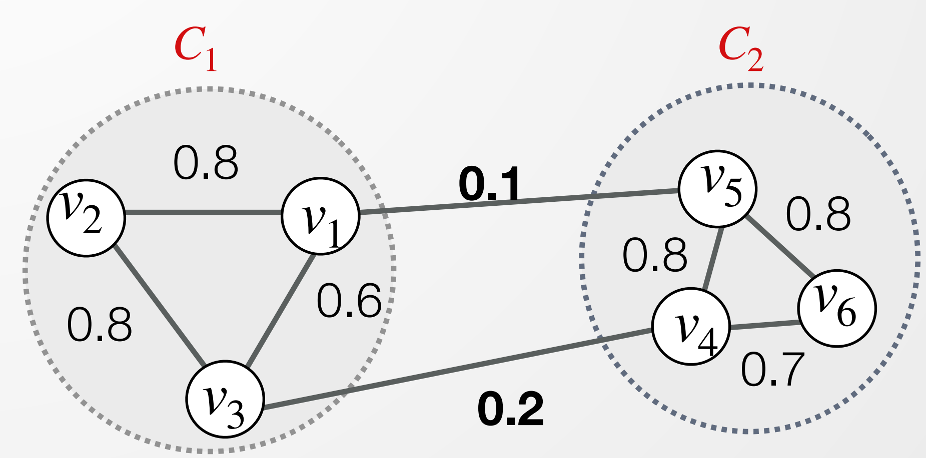
One connected component

Spectral Gap

Theoretical Example

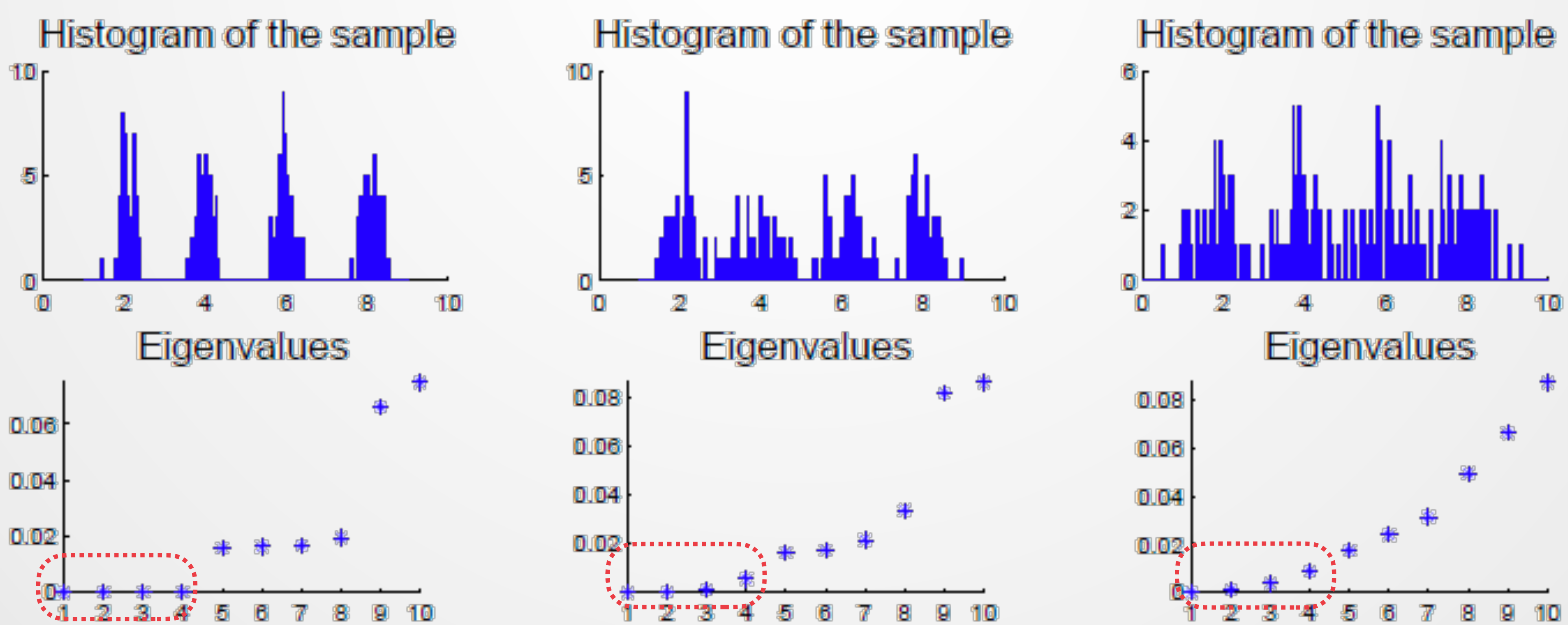
$$EVecs = \begin{bmatrix} -0.408 & -0.408 & 0.647 & 0.305 & -0.379 & -0.106 \\ -0.408 & -0.442 & -0.014 & -0.305 & 0.706 & -0.215 \\ -0.408 & -0.371 & -0.638 & -0.045 & -0.388 & 0.368 \\ -0.408 & 0.371 & -0.339 & 0.455 & -0.001 & -0.612 \\ -0.408 & 0.405 & 0.166 & 0.305 & 0.351 & 0.652 \\ -0.408 & 0.445 & 0.178 & -0.716 & -0.289 & -0.087 \end{bmatrix}$$





K-way Spectral Clustering

- Partitioning in *k* clusters
  - Take first *k* non-zero EVals
  - Build matrix $V \in \mathbb{R}^{n \times k}$ with EVecs as columns
  - Interpret rows of V as new data points, $Z_i \in \mathbb{R}^k$
  - Cluster the points Z_i with k-means algorithm in $\mathbb{R}^k$
- How many clusters?
  - Eigengap: We expect k EVals to be small and rest to be large
  - Example: 3 datasets (1-d, Gaussian with increasing variance)



Three basic stages

- Preprocessing
  - Construct a matrix representation of data
- Decomposition
  - Compute eigenvalues (EVal) and eigenvectors (EVec)
  - Map each point on lower dimensional representation based on one or more eigenvalues
- Grouping
  - Assign points to two or more clusters based on the new representation

Formalized Algorithm

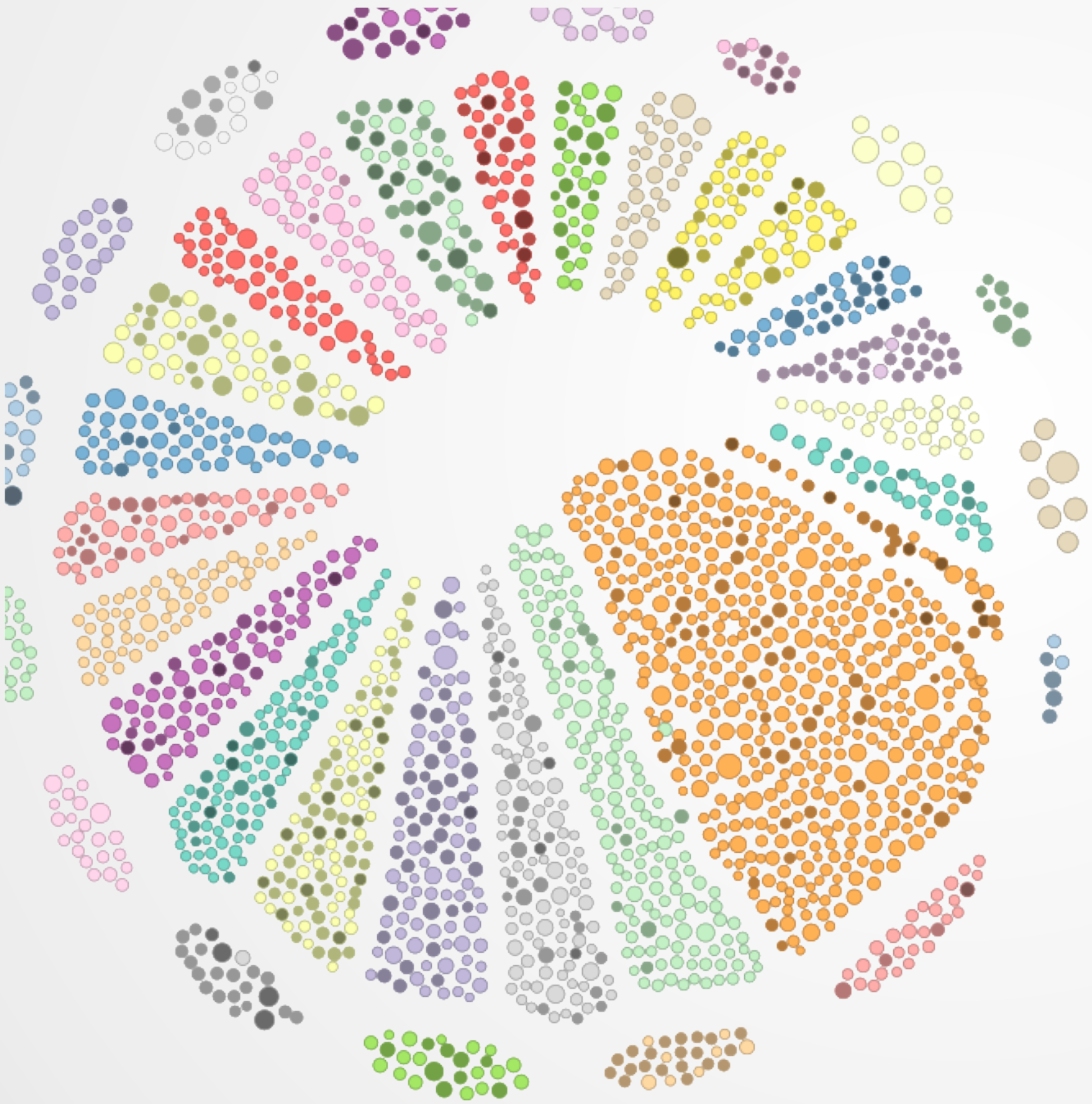
Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, k clusters to construct

- Construct a similarity graph by one of the ways described in Slide 6.
- Let W be its weighted adjacency matrix and D be the degree matrix
- Compute the Laplacian L.
- Compute the first k Evecs $v_1, \ldots, v_k$ of L.
- Let $V \in \mathbb{R}^{n \times k}$ be the matrix containing the Evecs $v_1, \ldots, v_k$ as columns.
- For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the *i*-th row on V.
- Cluster the points $(y_i)_{i=1,\ldots,n} \in \mathbb{R}^k$ with the *k*-means algorithm into clusters $C_1, \ldots, C_k$.

Output: Clusters $A_1, \ldots, A_k$ with $A = \{j \,|\, y_j \in C_j\}$.

Practical Example

- Spectral clustering of Quantlets using keywords



Quantlet Website

Quantlet Metainfo Example

Quantlet_Extraction_Evaluation_Visualisation

Name of QuantLet : Quantlet_Extraction_Evaluation_Visualisation

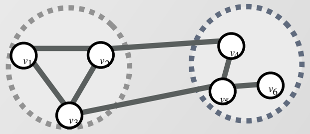
Published in : ''

Description : 'Extraction, grading and clustering of the Quantlets in the GitHub Organization Quantlet with

Keywords : Text analysis, LSA, t-SNE, clustering, kmeans clustering, spectral clustering, visualisation

See also : ''

Quantlet_Extraction_Evaluation_Visualisation



Data

"BCS/BCS_Binpdf/MetaInfo.txt": {"py/object": "modules.METAFILE.METAFI
"wagneran7@gmail.com", "name": "Palindroma"}, "comment_count": 0, "co
"Palindroma"}, "message": "MDI\n\nChapter 3 complete", "tree": {"sha"
"https://api.github.com/repos/QuantLet/BCS/git/trees/3dc1ce0c7513a666
"https://api.github.com/repos/QuantLet/BCS/git/commits/e5dc9c20a5ca99
"signature": null, "verified": false}}, "sha": "e5dc9c20a5ca99d6c303b
09T16:04:15Z", "email": "admin@quantnet.technology", "name": "Lukas B
"admin@quantnet.technology", "name": "Lukas Borke"}, "message": "auto
"https://api.github.com/repos/QuantLet/BCS/git/trees/a1b05133c04b2888
"https://api.github.com/repos/QuantLet/BCS/git/commits/701a6d4bcd2368
"signature": null, "verified": false}}, "sha": "701a6d4bcd23685790e2b
{"comment": "", "description_length": 43, "description_length_wo_stop
"submitted_year": "2018"}, "is_debuggable": true, "keyword_list": ["a
"visualization"], "last_modified": "Mon, 28 May 2018 16:09:32 GMT", "
distribution is plotted for sample sizes 5, 10 and 100 and probabilit
values and show when it can likely be approximated by the normal dist
representation, pdf, visualization", "name of quantlet": "BCS_Binpdf"

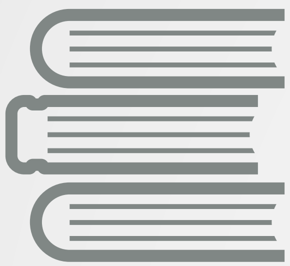
▾ root: {} 8 keys
py/object: "modules.QUANTLET.QUANTLET"
▸ errors: [] 5 items
▸ g: {} 44 keys
github_token: null
▸ keywords_stats: {} 2 keys
▸ last_full_check: {} 2 keys
▸ quantlets: {} 2124 keys
▸ repos: {} 166 keys

- Original Json File
- Data Tree Structure

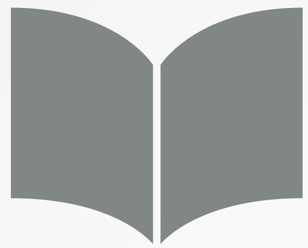
Data Source

Preprocessing example

Documents



Dictionary of Words



Remove stop words

cat sat mat

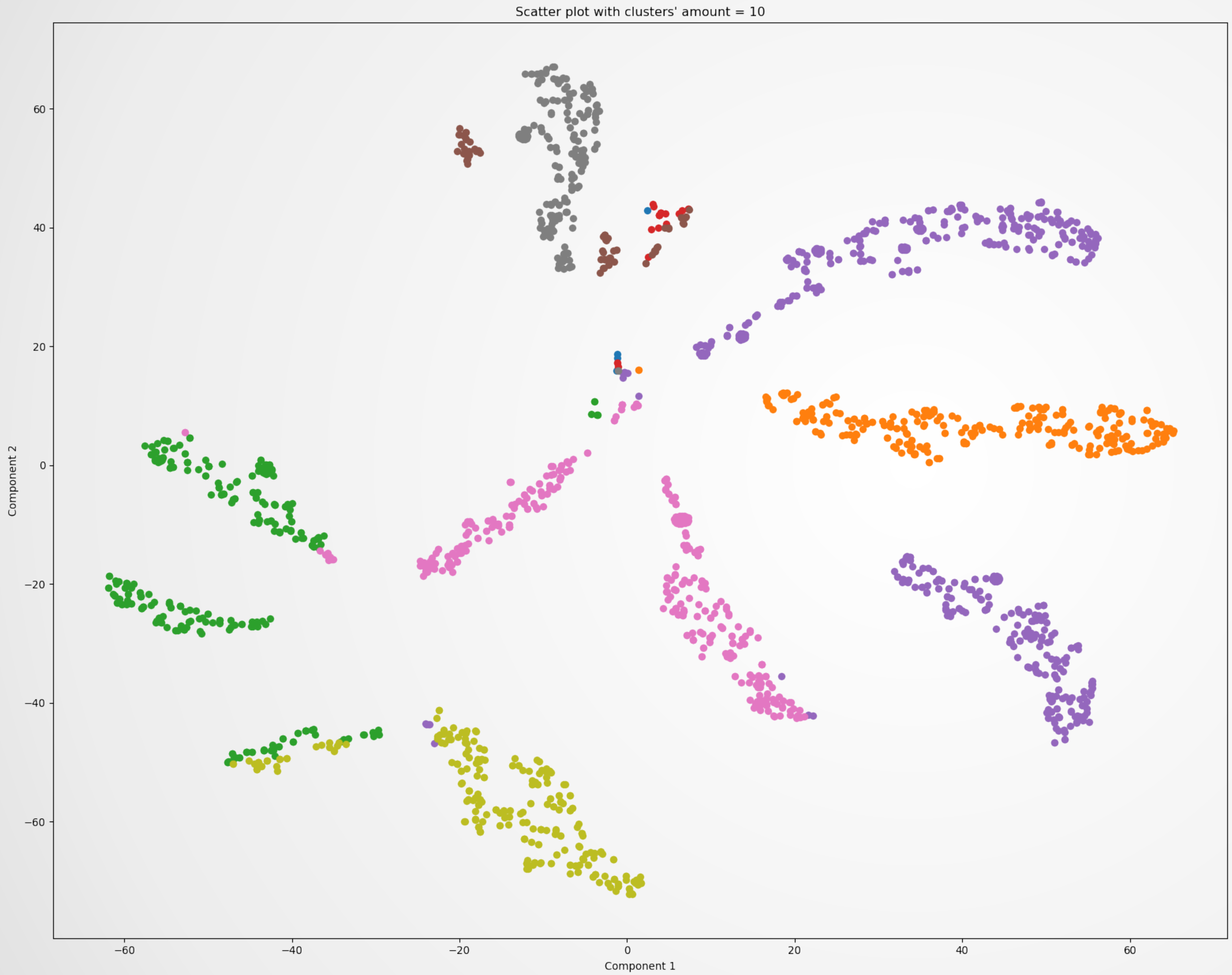
Example sentences:
„the cat sat on a mat“
„book from the shelf“

1. the
2. cat
3. sat
4. on
5. a
6. mat
7. book
8. from
9. the
10. shelf

Tokenize to a Tensor

	cat	sat	mat	book	shelf
the cat sat on a mat	1	1	1	0	0
book from the shelf	0	0	0	1	1

Visualization t-SNE



- 'work not did **clustering similarity** lstm **kernel density** predictions'
- 'plot graphical representation visualization distribution time analysis series returns'
- '**plot graphical representation visualization** data simulation financial time regression'
- 'plot var networks mandatory'
- 'Pareto density **crix** distribution volatility plot stable **crixcoin cryptocurrency**'
- 'plot graphical representation data visualization estimation series time financial'
- 'plot graphical representation visualization data financial time series distribution'
- 'plot visualization simulation distribution graphical time representation series regression'
- 'plot distribution estimation financial regression density quantile prediction simulation'

Classification example (Cluster 5)

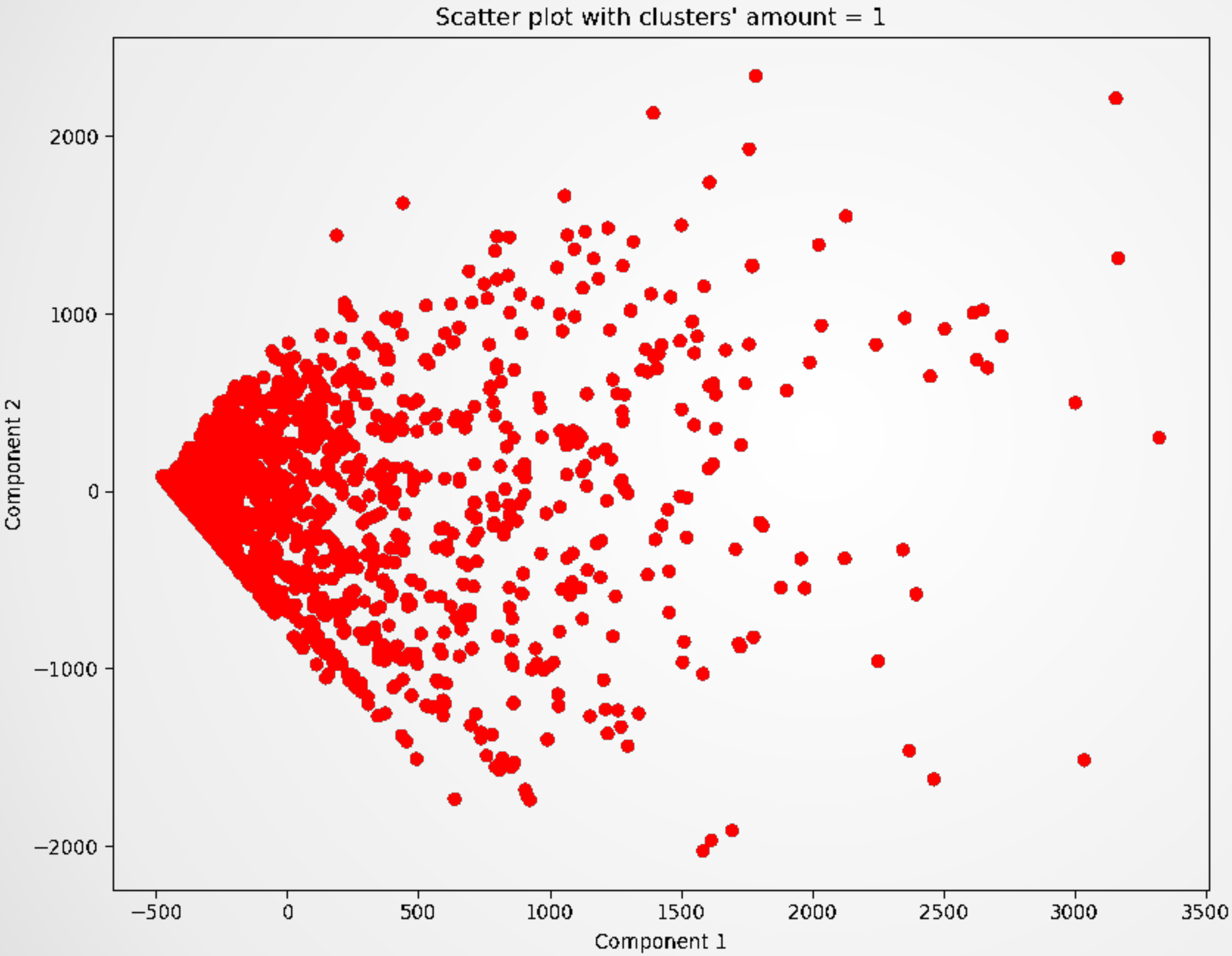
By topic (CCs)

237	CRIX/CRIXind
244	CRIX/FDAXmembers
246	CRIX/FIPCmembers
248	CRIX_tedas/CRIXcoreWealth
253	CRIX_tedas/CRIXtedasCorrelationCoreCryptos
286	CryptoDynamics/CryptoDynamics_Estimation/m
288	CryptoDynamics/CryptoDynamics_Series/m
289	CryptoDynamics/CryptoDynamics_Wachter/m
302	DEDA-Bitcoin/Bitcoin_Evalu
303	DEDA-Bitcoin/Bitcoin_Simul

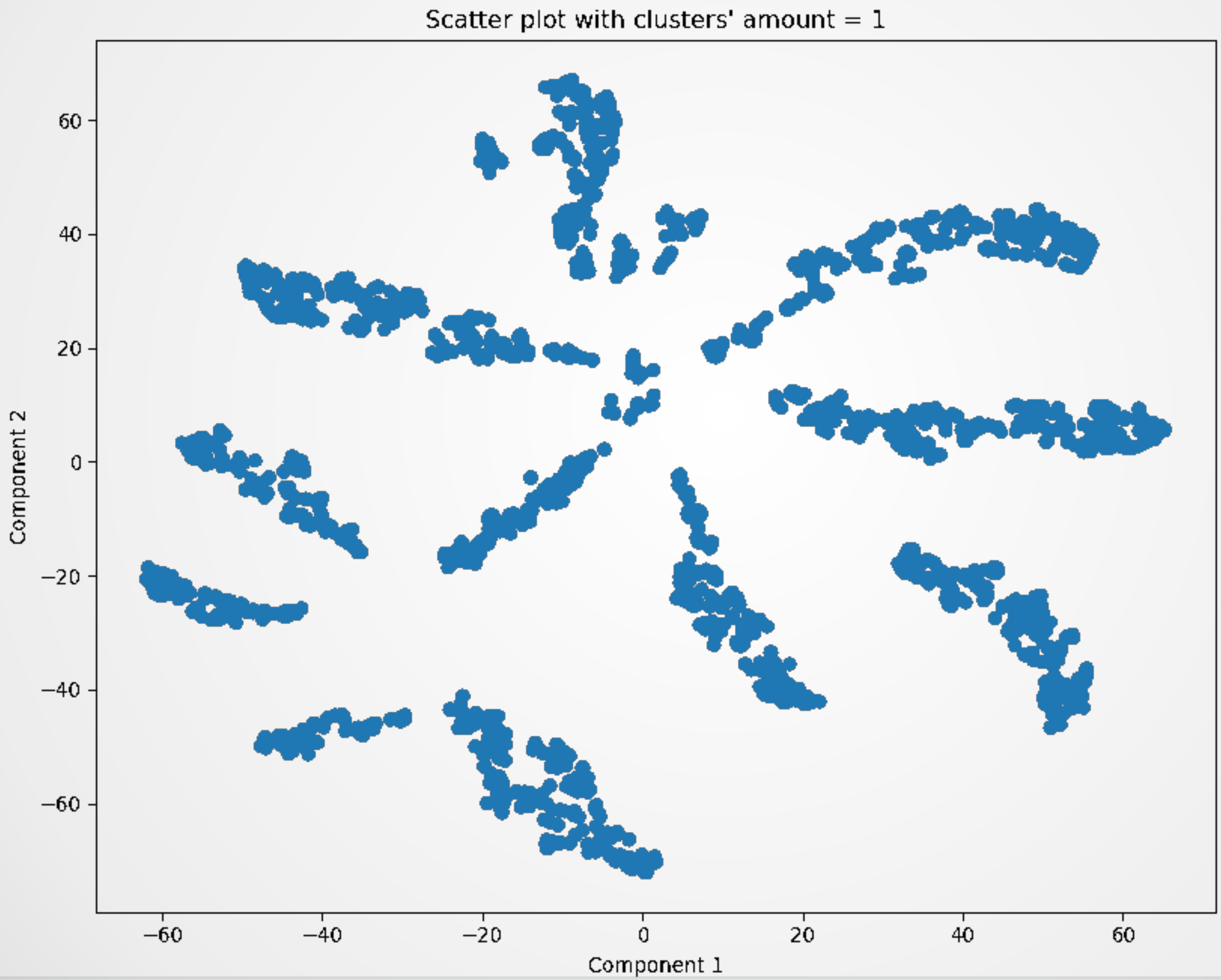
By user

1070	NextUnicorn/NextUnicorn_AUC
1071	NextUnicorn/NextUnicorn_ConditionalTr
1072	NextUnicorn/NextUnicorn_DataCleaning
1073	NextUnicorn/NextUnicorn_DescriptiveStats
1074	NextUnicorn/NextUnicorn_ExitDeal
1075	NextUnicorn/NextUnicorn_LogisticRegress
1076	NextUnicorn/NextUnicorn_RandomFores
1077	NextUnicorn/NextUnicorn_RecursivePartitioning
1078	NextUnicorn/NextUnicorn_Results
1079	NextUnicorn/NextUnicorn_Scatter
1080	NextUnicorn/NextUnicorn_Sigmoid
1081	NextUnicorn/NextUnicorn_XGBoos
1082	NextUnicorn/NextUnicorn_helperFunctions

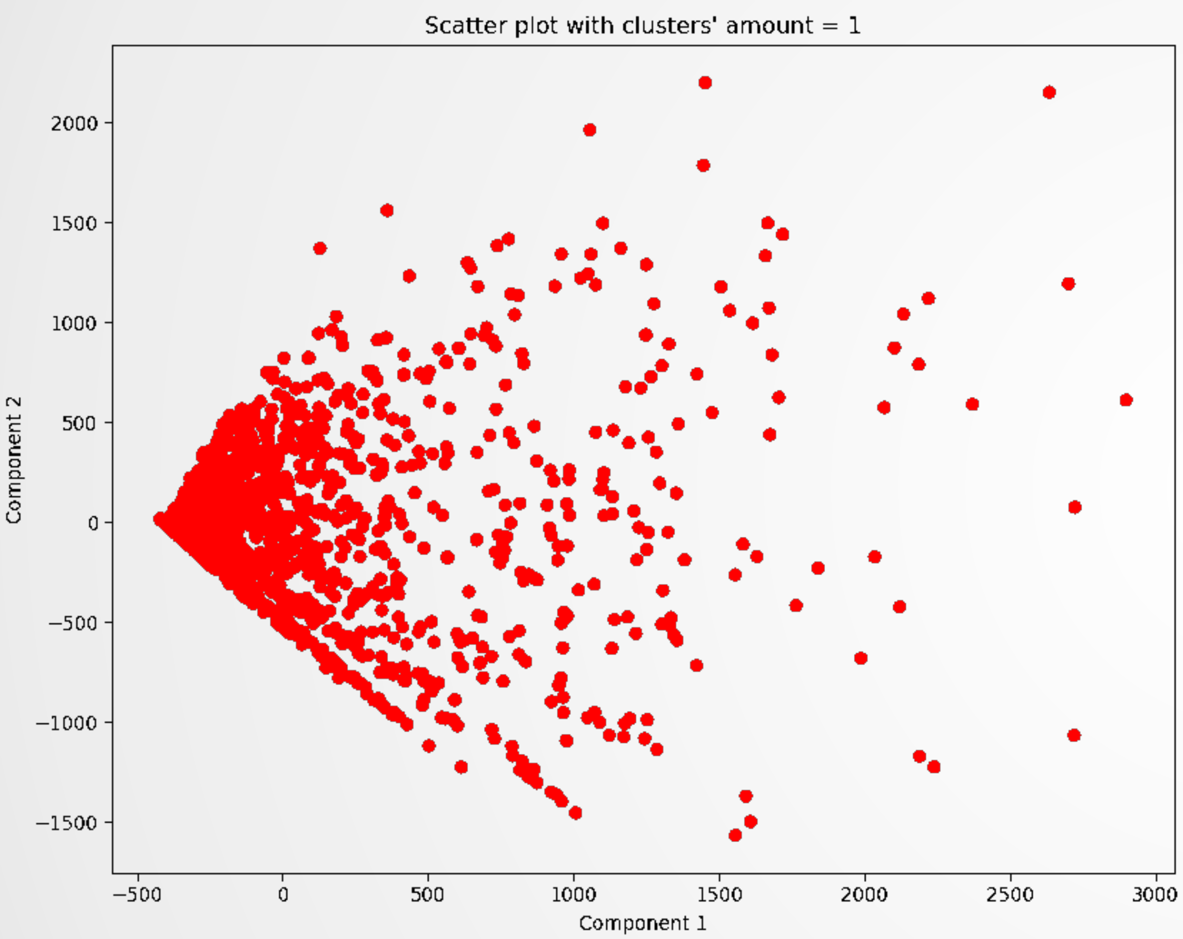
Dynamics of different amount of clusters (PCA)



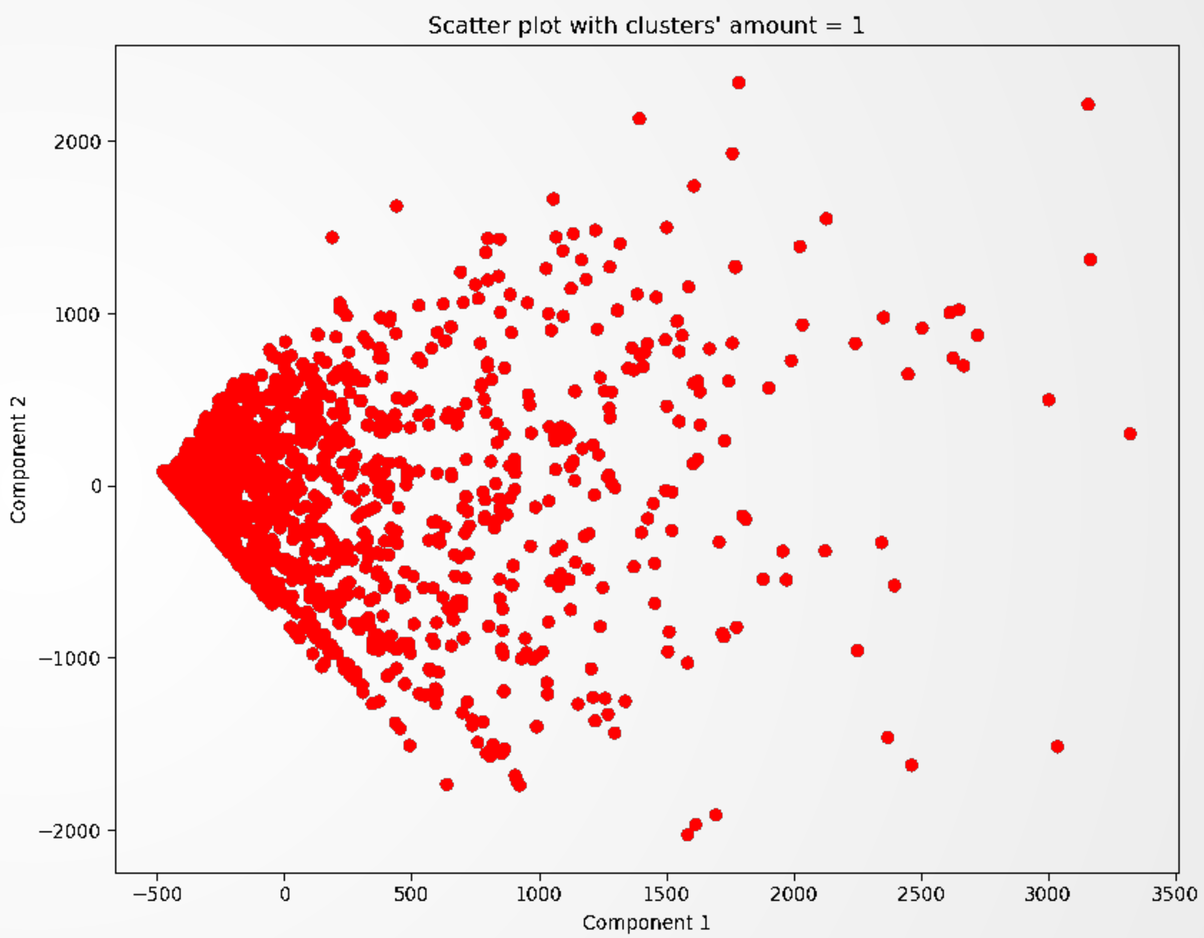
Dynamics of different amount of clusters (t-SNE)



Hierarchical Clustering



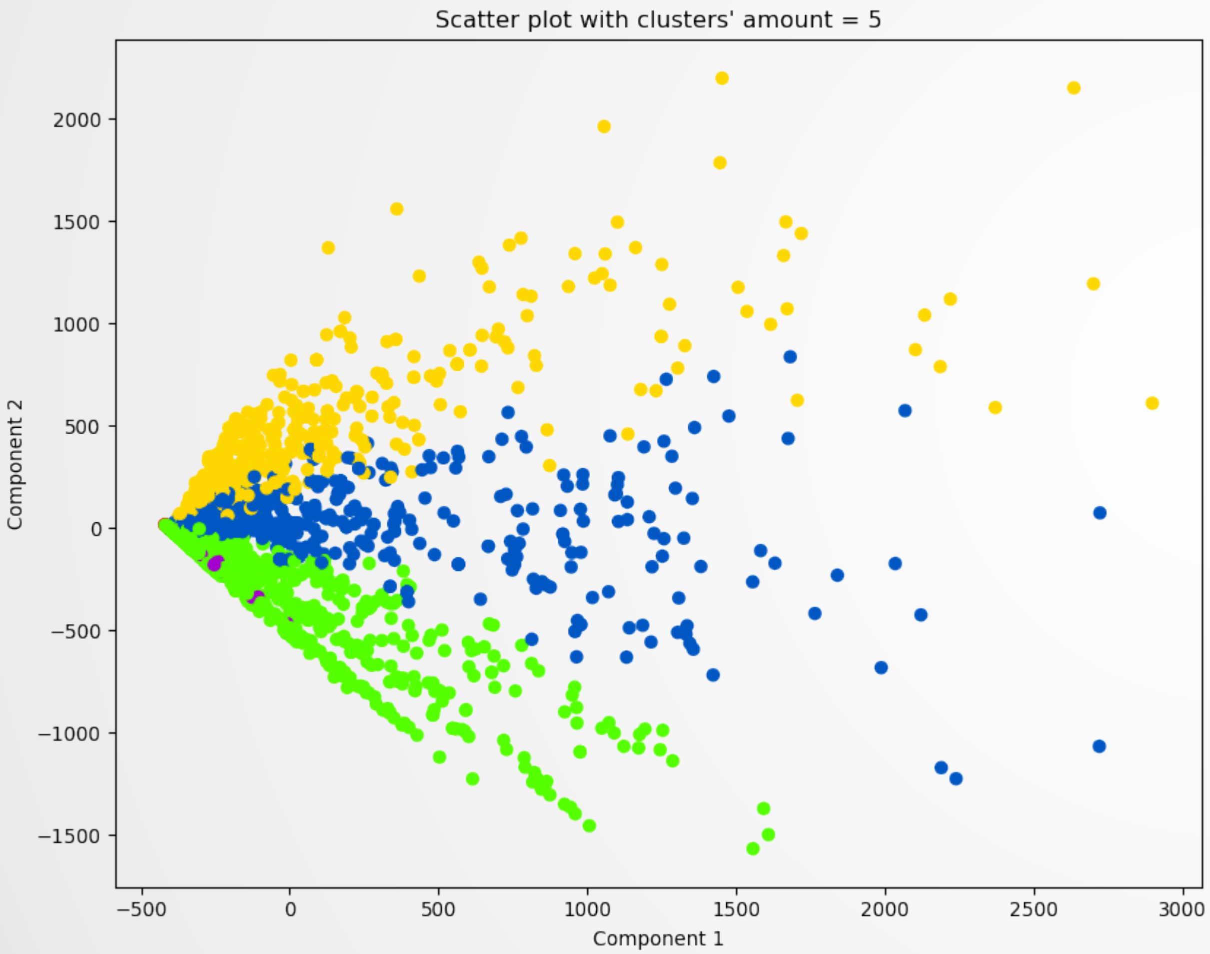
Spectral Clustering



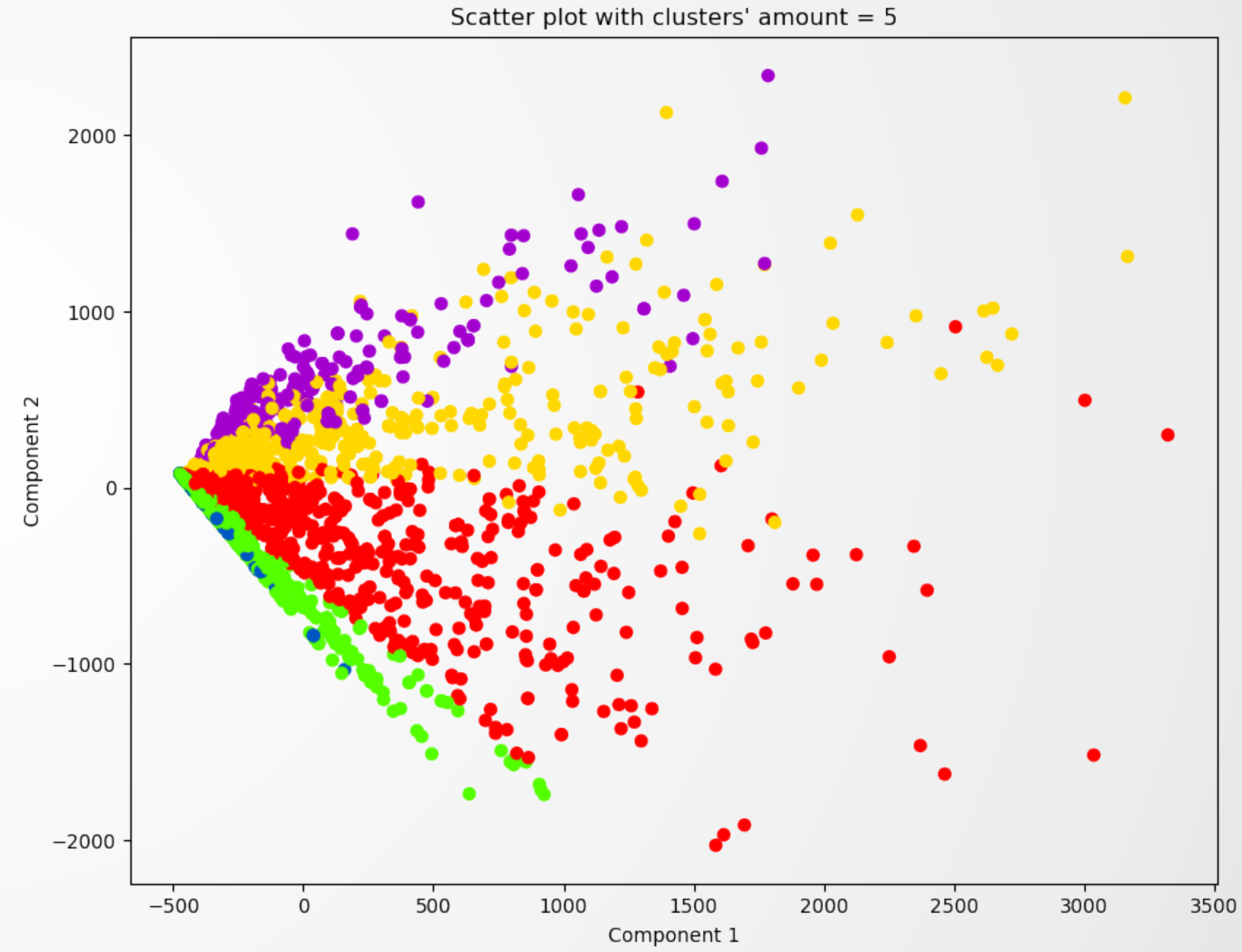
Note: More balanced clusters with SC

Number of clusters = 5

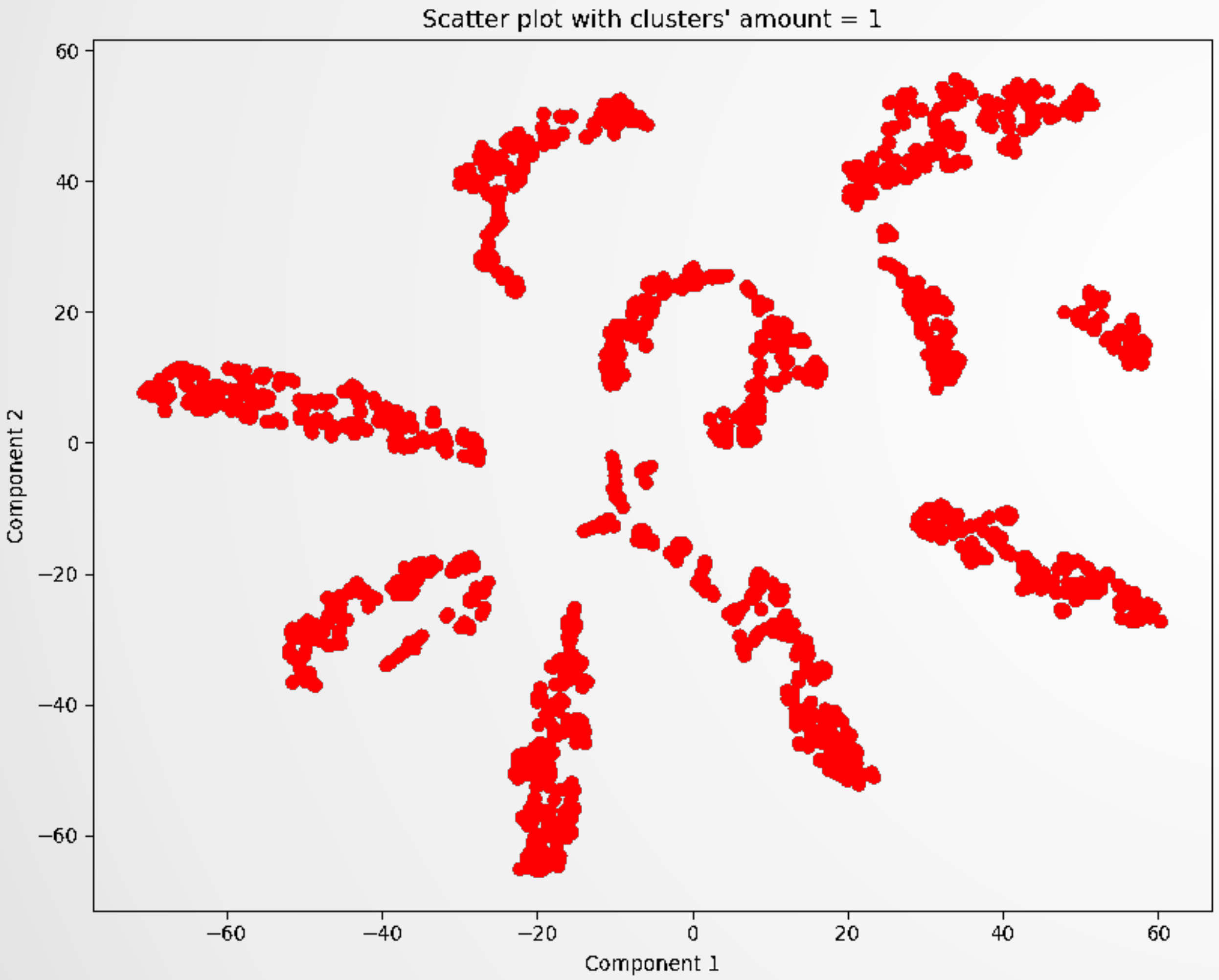
Hierarchical Clustering



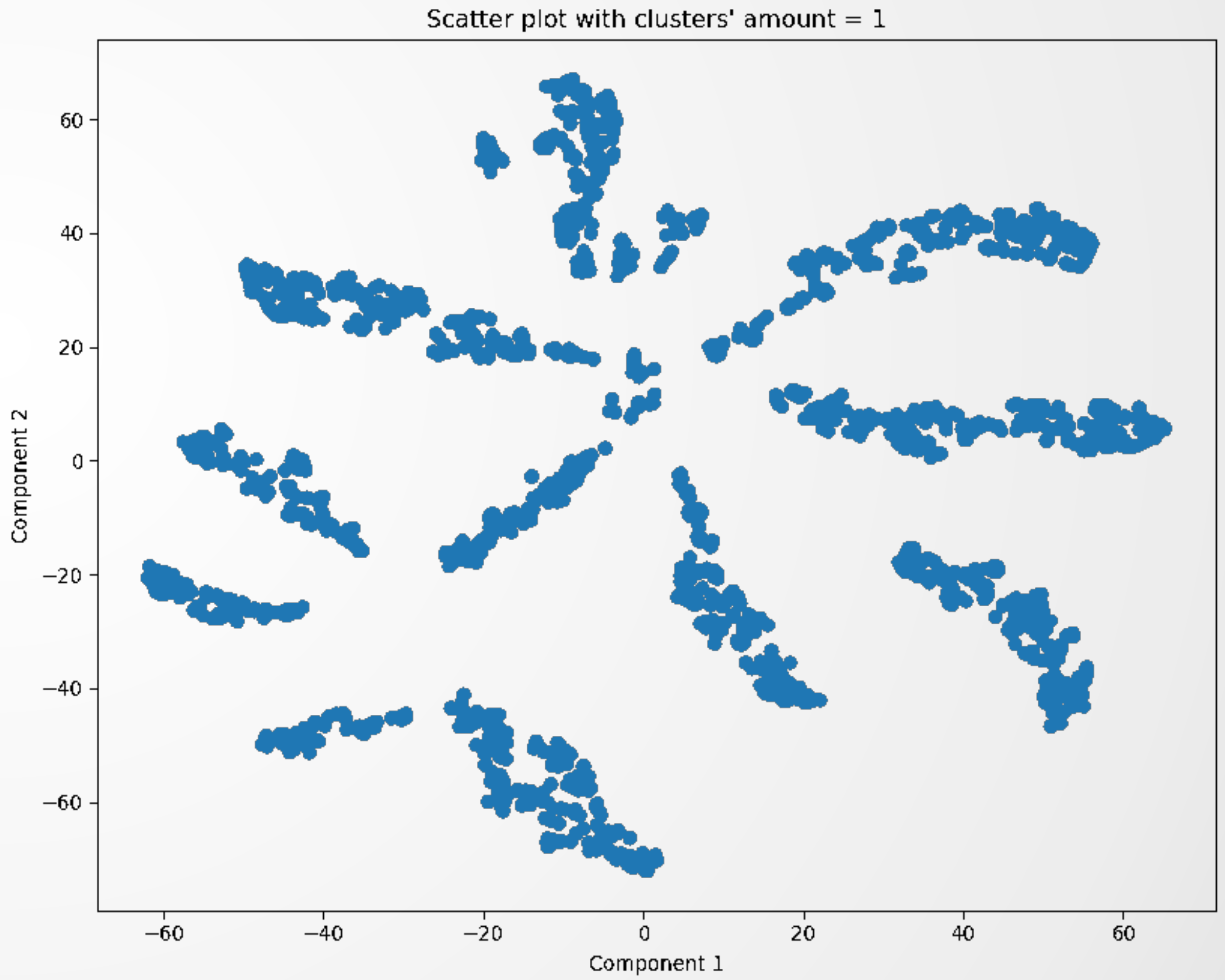
Spectral Clustering



Hierarchical Clustering



Spectral Clustering



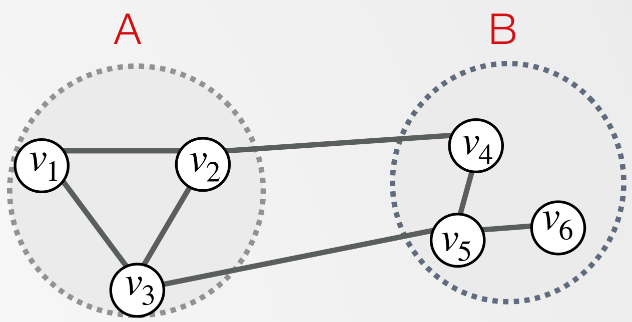
Takeaways

- Success of SC is mainly based on the fact that it does not make any assumptions on the form of the clusters.
- SC can solve very general problems like intertwined spirals (non-convex sets)
- SC can be implemented efficiently even for large data sets, as long as we make sure that the similarity graph is sparse (no issue of getting stuck at local minima, just linear problem)
- Choosing a similarity graph can be tricky
- Sc is a powerful tool for clustering but not a black box to determine how many clusters to use.

References

- https://drive.google.com/file/d/0B1InOG_uMMAnRGZJSk5ZczNHUUU/view
- https://ai.stanford.edu/~ang/papers/nips01-spectral.pdf
- http://www.tml.cs.uni-tuebingen.de/team/luxburg/publications/Luxburg07_tutorial.pdf
- http://evoq-eval.siam.org/Portals/0/Publications/SIURO/Vol5/SPECTRAL_CLUSTERING.pdf?ver=2018-04-06-121714-460
- https://towardsdatascience.com/spectral-clustering-aba2640c0d5b
- slideshare.net/hnly228078/spectral-clustering-tutorial
- https://www.datasciencecentral.com/profiles/blogs/comparing-spectral-partitioning-clustering-with-normalized-graph
- Li G, Tubo Y, Härdle WK (2019) Dynamic Network Perspective of Cryptos, J Amer Stat submitted.
- Iyigun C. et al 2013
- Cai X. et al 2011
- Bian J. et al 2019
- Koch Institute Gallary
- Mohammad S & Salama M 2007

Spectral Clustering



Kainat Khowaja

Wolfgang Karl Härdle

IRTG 1792
High Dimensional Non-Stationary Time Series
Humboldt-Universität zu Berlin
IRTG1792.HU-Berlin.de