Hyperbolic K-means for Cloud- and virtualised-RANs

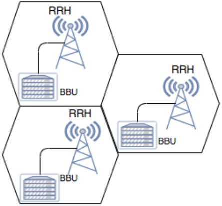
Anastasios Giovanidis
with H. Djeddal, L. Touzari, C.-D. Phung & S. Secci

Cooperation: Sorbonne University CNRS-LIP6 & CNAM Cédric

31/08/2021

Traditional RAN

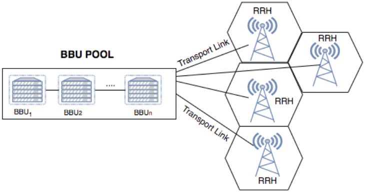
- 5G : massively connected objects (IoT, vehicles, monitoring, etc.)
- Increased demand for edge-computing and virtualisation



☞ On each site: Remote Radio Head (RRH) and Base-Band Unit (BBU).

Cloud-RAN

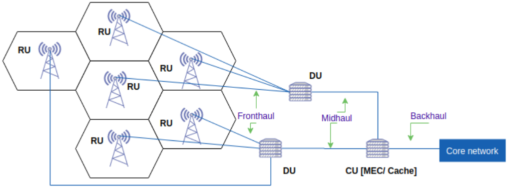
- Low deployment (CapEx) with energy savings (OpEx).
- When a station is idling, resources can be shared by another.
- enables cooperative communications (CoMP).



☞ BBUs centralised in a pool allowing dynamic resource allocation.

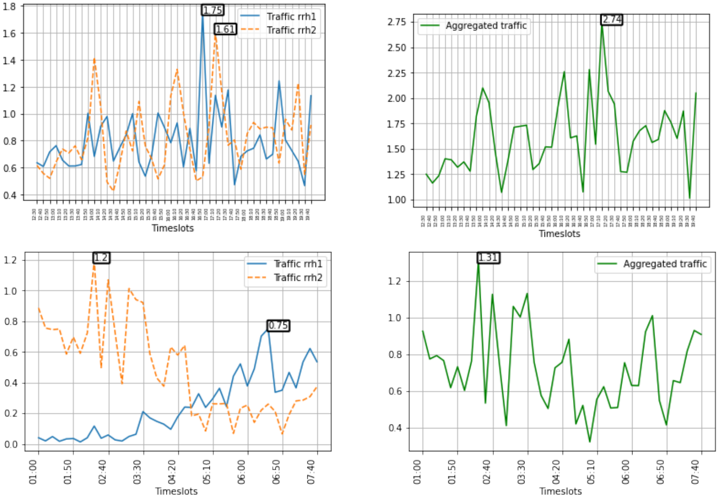
virtualised-RAN

- BBU controls run as software on generic hardware (virtualisation).
- Resource allocation and load balancing adapted on demand.
- C-RAN can be seen as an application running on vRAN.



☞ Radio Units (RUs) for coverage, Distributed Units (DUs) for layer-1 and -2 functions, Centralised Unites (CUs) for Base-band operations.

Resource economy from traffic aggregation



Clustering problem

How many clusters to form?
Which RUs to cluster together in a common CU?

It depends on service delay requirements and fibre link capacity.

Clustering problem

How many clusters to form?
Which RUs to cluster together in a common CU?

It depends on service delay requirements and fibre link capacity.

- Typical macro-cell can support up to 15 RUs *[Pizzinat et al 2015]* (Orange Labs Lannion).
- Aggregate all RUs of a small town to a unique BBU pool.
- Round-Trip-Time (RTT) budget is 3 *ms* due to H-ARQ delay requirements (corresponds to 20 *km* fronthaul length).
- For time-critical uRLLC applications, RTT budget drops to 1 *ms* (so fronthaul length should be much shorter).

A meaningful RU clustering

- Reduce BBU processing capacity (by maximising BBU utilisation).
- Fair traffic load balancing among clusters.
- Respect optical fibre length limitations.
- Form geographically compact clusters to avoid frequent inter-cluster handovers.

Select a number $K \geq 1$; find a meaningful partition of the RUs, given their 2D planar position and traffic profile over time

$$\mathcal{P} = \{\mathcal{C}_1, \ldots, \mathcal{C}_K\}.$$

Basic traffic quantities

RU stations: $n = 1, \ldots, N$ with fixed 2D-position (x_n, y_n).
Time slots: $t = 1, \ldots, T$. Time-slot duration e.g. 1 *hour*, 10 *mins*...

- Per station traffic
  - Normalised measured traffic for station n and slot t:

$$z(n, t) \in [0, 1].$$

  - (*) Remaining available resources for station n and slot t:

$$w(n, t) = 1 - z(n, t)$$

- A group $\mathcal{C}$ of stations:
  aggregate traffic per time slot $Z(\mathcal{C}, t) = \sum_{n \in \mathcal{C}} z(n, t)$, for each t.
  (*) aggregate rest resources per slot $W(\mathcal{C}, t) = |\mathcal{C}| - \sum_{n \in \mathcal{C}} z(n, t)$.

Existing methods

- Mixed Integer Linear Programming (MILP) *[da Silva Coelho et al 2020]*: NP-hard; Heuristic solutions for large instances; Sensitivity of solutions to choice of costs and parameters.
- Distance-Constrained Clustering Algorithm (DCCA) *[Chen et al 2018]*: considers both 2D position and traffic-demand; Groups together RUs with traffic-complementarity, based on an entropy measure; Greedy algorithm, no guarantees.
- K-means in 2D: uses the 2D RU positions as input; Does not include traffic information; Each cluster groups neighbouring RUs around a (chosen) centroid.

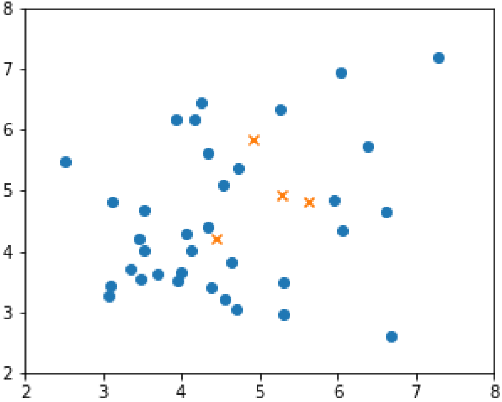
K-means in 2D Euclidean

☞ Minimises the within cluster variance

$$\min_{\mathcal{P}} \sum_{k=1}^{K} Var(\mathcal{C}_k), \qquad Var(\mathcal{C}_k) = \frac{1}{|\mathcal{C}_k|} \sum_{i \in \mathcal{C}_k} (d_{E2}(i, m_k))^2$$

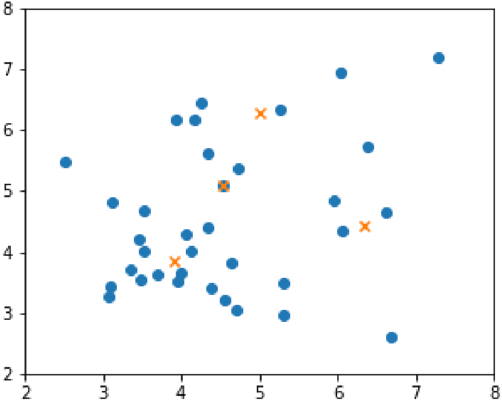
- Initialisation: place randomly the K centroids $m_1, \ldots, m_K$ on 2D.
- Iterate:
  1. Assign each RU to the $2D$-Euclidean closest centroid.
  2. Update all centroid coordinates $m_k = (\overline{x}_k, \overline{y}_k)$ as the arithmetic mean of the associated RU 2D-coordinates.
- Convergence: till the relative change of all centroids falls below some threshold.

K-means in 2D Euclidean



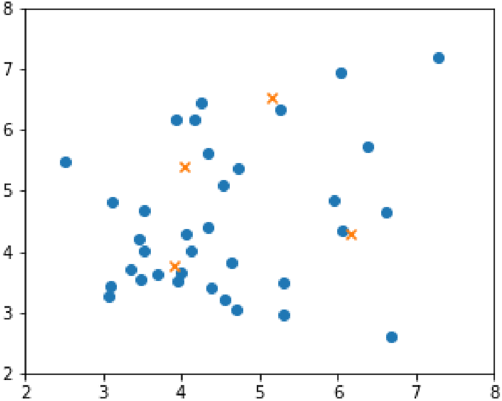
Initialisation

K-means in 2D Euclidean



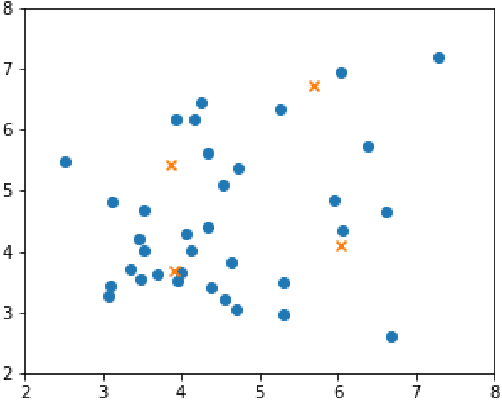
Update 1

K-means in 2D Euclidean



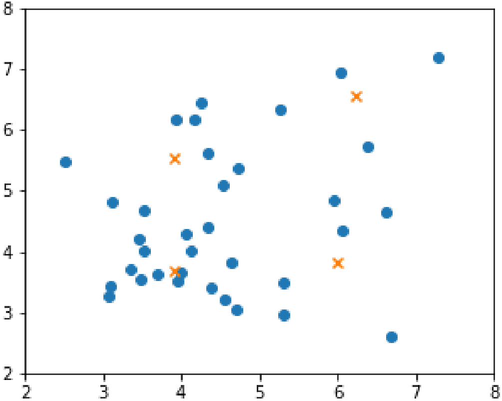
Update 2

K-means in 2D Euclidean



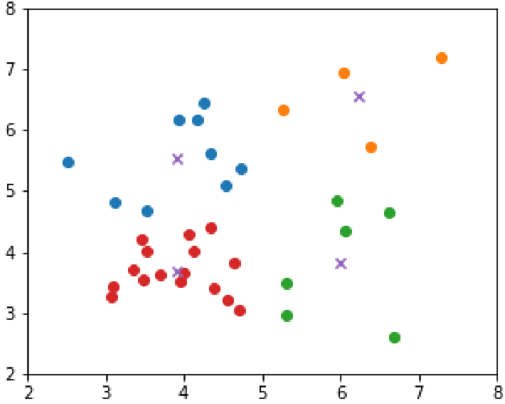
Update 3

K-means in 2D Euclidean



Update 4

K-means in 2D Euclidean



Aggr. Traffic $[Z_1,\ Z_2,\ Z_3,\ Z_4] = [4.1,\ 1.8,\ 4.1,\ 7.7]$, $Var_Z = 5.9$

K-means in 3D Euclidean

- Data points

$$(x_i, y_i) \rightarrow (x_i, y_i, z_i) \text{ or } w_i.$$

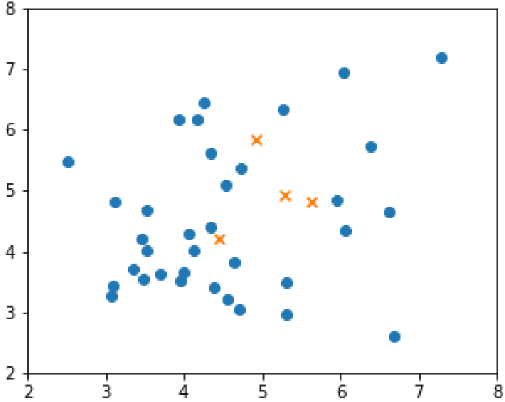
- Distance

$$d_{E2}((x_i, y_i), (x_j, y_j)) \rightarrow d_{E3}((x_i, y_i, z_i), (x_j, y_j, z_j))$$

Notice the relations:

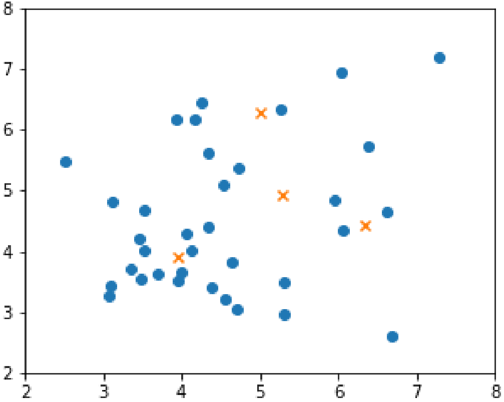
- $d_{E3}^2((x_i, y_i, z_i), (x_j, y_j, z_j)) = d_{E2}^2((x_i, y_i), (x_j, y_j)) + (z_i - z_j)^2$
- $d_{E3}^2((x_i, y_i, z_i), (x_j, y_j, z_j)) = d_{E3}^2((x_i, y_i, w_i), (x_j, y_j, w_j)).$

K-means in 3D Euclidean



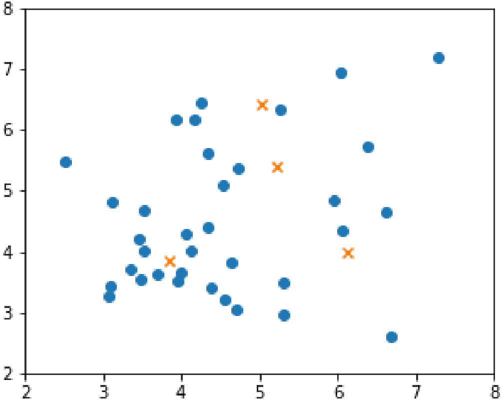
Initialisation

K-means in 3D Euclidean



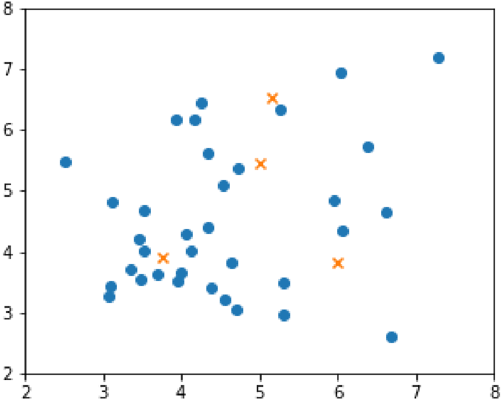
Update 1

K-means in 3D Euclidean



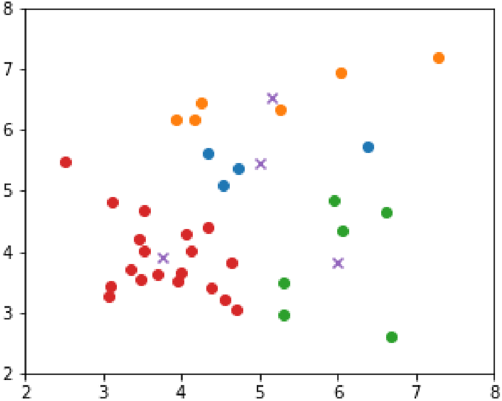
Update 2

K-means in 3D Euclidean



Update 3

K-means in 3D Euclidean



Aggr. Traffic $[Z_1,\ Z_2,\ Z_3,\ Z_4] = [1.9,\ 2.8,\ 4.1,\ 8.9]$, $Var_Z = 9.7$

K-means in 3D Euclidean

- The clusters are misbalanced in traffic.
- They are not very compact (spread in one dimension).

☞ Why doesn't it work?

$d_{E3}^2 = d_{E2}^2 + (z_i - z_j)^2$: when traffic is equal RUs appear close

☞ Clusters RUs with equal rather than unequal traffic!

3D Hyperbolic distance

☞ Distance

$$d_{H3}((x_i, y_i, w_i), (x_j, y_j, w_j)) = arccosh\left(1 + \frac{{d_{E3}}^2}{2w_i w_j}\right)$$

with $arccosh(s) = ln(s + \sqrt{s^2 + 1})$.

Properties: positivity, identity of indiscernibles, symmetry, sub-additivity.

Hyperbolic Metric Space

☞ Data is assumed embedded on the Poincaré half-plane model

$$\hat{\mathbb{B}}^3 \;=\; \{(x,y) \in \mathbb{R}^2,\; w \in \mathbb{R}_+ \;:\; ||(x,y,w)|| < 1\}$$

- The set is equipped with distance d_{H3} to form a **metric space**.

Hyperbolic Metric Space

☞ Data is assumed embedded on the Poincaré half-plane model

$$\hat{\mathbb{B}}^3 \;=\; \left\{(x,y) \in \mathbb{R}^2,\; w \in \mathbb{R}_+ \;:\; ||(x,y,w)|| < 1\right\}$$

- The set is equipped with distance d_{H3} to form a **metric space**.

☞ Hyperbolic learning has been very successful recently for:

- the analysis of complex networks [Krioukov et al 2010]
- learning data with hierarchy and similarity [Nickel and Kiela 2017]
- community detection in graphs [Hajri et al 2019]

Through the Hyperbolic Lens...

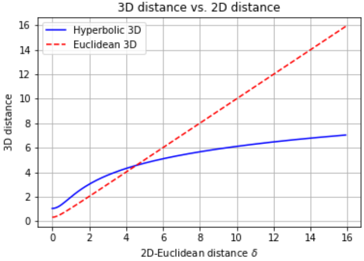
$$d_{H3}((x_i, y_i, w_i), (x_j, y_j, w_j)) = arccosh\left(1 + \frac{{d_{E3}}^2}{2 w_i w_j}\right)$$

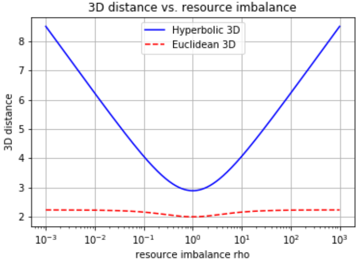
- When $w_i w_j$ is low, RUs appear distant.
- Consequence: 2 RUs both with low resources (high load) are placed in separate clusters.

Through the Hyperbolic Lens...

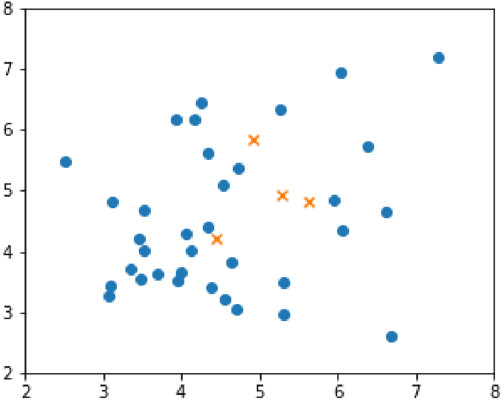
$$d_{H3}((x_i, y_i, w_i), (x_j, y_j, w_j)) = arccosh\left(1 + \frac{{d_{E3}}^2}{2 w_i w_j}\right)$$

- When $w_i w_j$ is low, RUs appear distant.
- Consequence: 2 RUs both with low resources (high load) are placed in separate clusters.



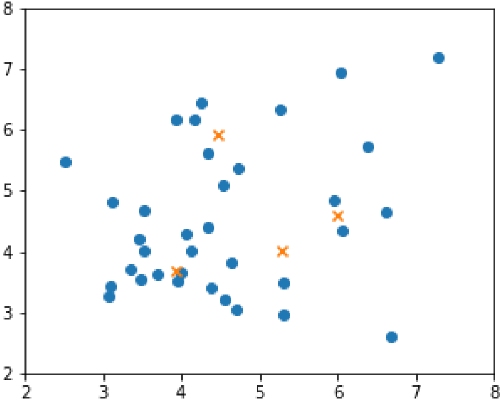


K-means in 3D Hyperbolic



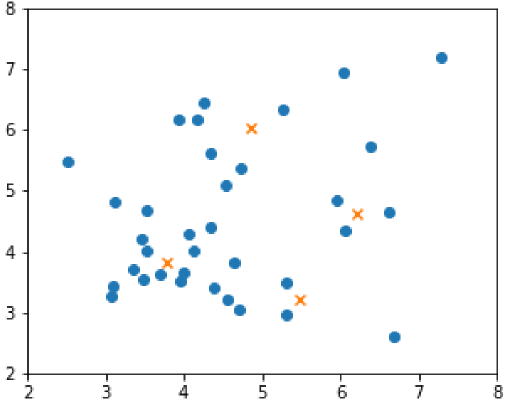
Initialisation

K-means in 3D Hyperbolic



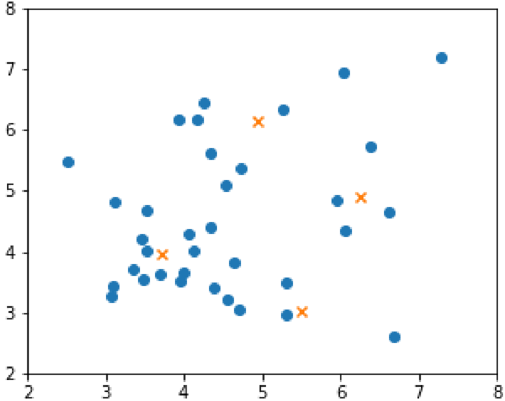
Update 1

K-means in 3D Hyperbolic



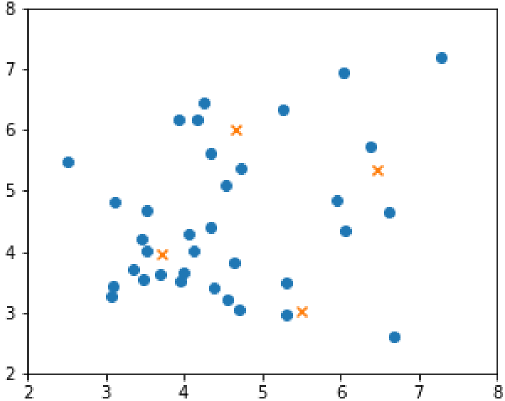
Update 2

K-means in 3D Hyperbolic



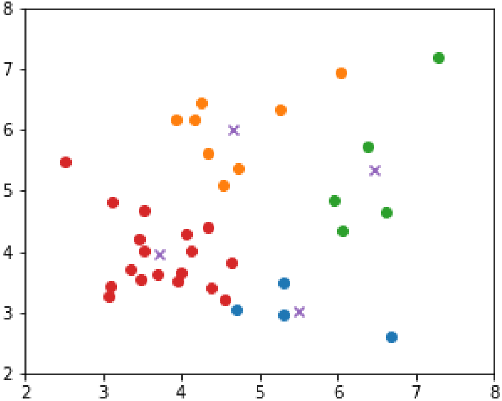
Update 3

K-means in 3D Hyperbolic



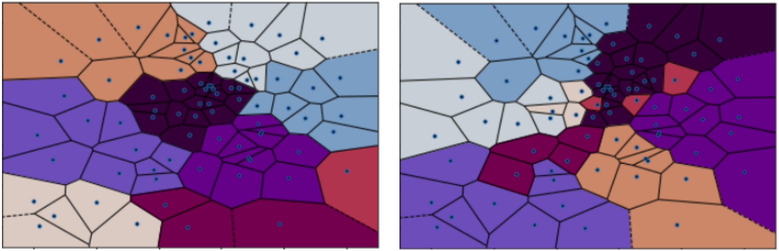
Update 4

K-means in 3D Hyperbolic



Aggr. Traffic $[Z_1, Z_2, Z_3, Z_4] = [3.1, 3.8, 2.8, 7.9]$, $Var_Z = 5.62$

Lille city (Orange data)



(a) 2D-euclidean, (b) 3D-hyperbolic

Temporal clustering

☞ K-means does not include the temporal change of traffic. How to?

Step.1 For each slot $t = 1, \ldots, T$:

- Execute `K-means in 3D Hyperbolic`.
  Obtain the set of centroids $\{m_1(t), \ldots, m_K(t)\}$.

Step.2 Given the K centroids for slot $t = 1$, get for each one the $T - 1$ closest (d_{E2}) centroids from every other partition $t = 2, \ldots, T$.

Step.3 Compute the average K centroids: each $(\overline{x}_k, \overline{y}_k, \overline{w}_k)$ is an average over T realisations.

Step.4 Assign each RU anew to the closest average centroid based on the hyperbolic (d_{H3}) distance.

Average-to-Peak traffic ratio

Average-to-Peak Traffic Ratio (AtPTR) per group $\mathcal{C}$

$$U(\mathcal{C}) = \frac{\frac{1}{T}\sum_{t=1}^{T} Z(\mathcal{C}, t)}{\max_t Z(\mathcal{C}, t)} \leq 1.$$

☞ The closer to 1, the better utilisation of the BBU resource over time.

Performance metrics

1. Utilisation for partition $\mathcal{P}$ *[Chen et al 2018]*

$$Util(\mathcal{P}) = \frac{\frac{1}{K}\sum_{k=1}^{K} U(\mathcal{C}_k)}{\frac{1}{N}\sum_{n=1}^{N} U(n)} \geq 1,$$

is a measure of improvement of the average AtPTR due to $\mathcal{P}$.
☞ $Util(\mathcal{P}) = 1$ for $K = N$. Maximised for $K = 1$.

Performance metrics

1. Utilisation for partition $\mathcal{P}$ *[Chen et al 2018]*

$$Util(\mathcal{P}) = \frac{\frac{1}{K}\sum_{k=1}^{K} U(\mathcal{C}_k)}{\frac{1}{N}\sum_{n=1}^{N} U(n)} \geq 1,$$

is a measure of improvement of the average AtPTR due to $\mathcal{P}$.
☞ $Util(\mathcal{P}) = 1$ for $K = N$. Maximised for $K = 1$.

2. Cost for partition $\mathcal{P}$ *[Chen et al 2018]*

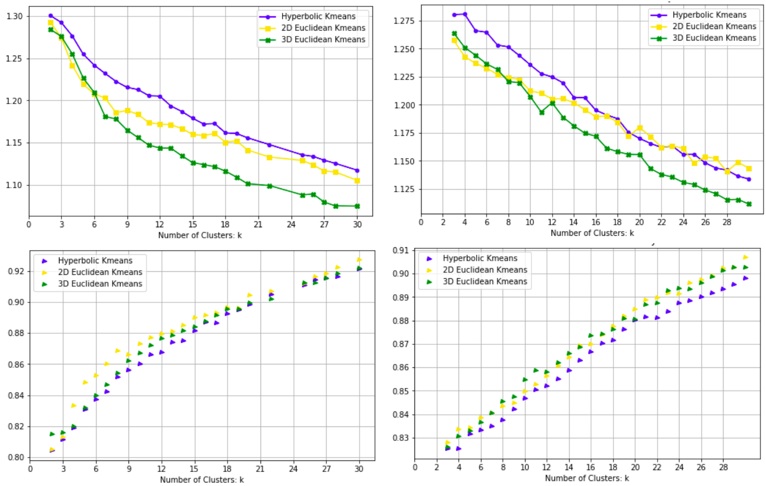
$$Cost(\mathcal{P}) = \frac{\sum_{k=1}^{K} \max_t Z(\mathcal{C}_k, t)}{\sum_{n=1}^{N} \max_t z(n, t)} \leq 1.$$

is a measure of cost reduction, due to $\mathcal{P}$.
☞ $Cost(\mathcal{P}) = 1$, for $K = N$. Minimised for $K = 1$.

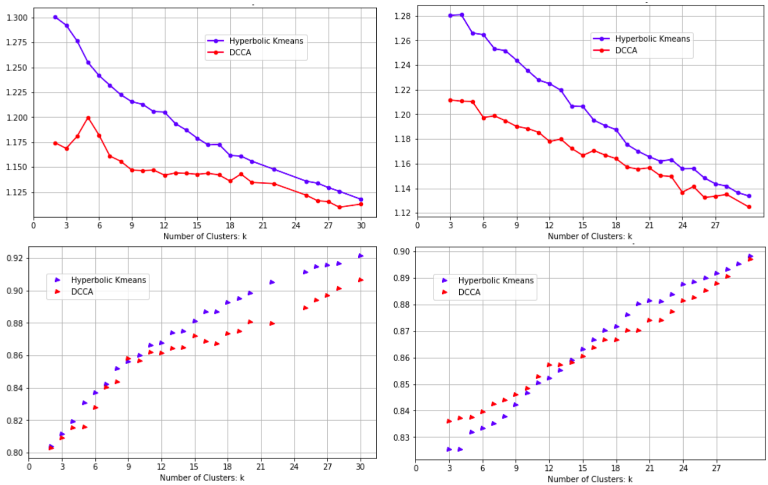
Data set

Dataset	Lille	Nantes
Number of RU positions	88	97
Data collection period	2019-03-19 to 2019-06-16	2019-03-16 to 2019-06-16
Time-slot duration	10-minute	10-minute
Maximal traffic Bytes-Up	$7.25\ 10^9$	$3.04\ 10^9$
Minimal traffic Bytes-Up	0.0	0.0
Maximal traffic Bytes-down	$15.82\ 10^9$	$18.15\ 10^9$
Minimal traffic Bytes-down	0.0	0.0

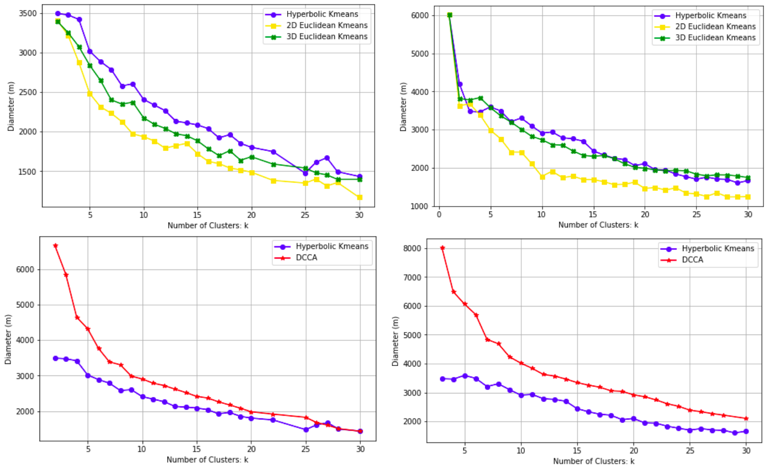
Hyperbolic VS Euclidean



Hyperbolic VS DCCA [Chen et al 2018]



Cluster Diameter



Lille and Nantes

END

Contact:
fh.djeddal@gmail.com
anastasios.giovanidis@lip6.fr

Code:
https://github.com/hanane-djeddal/Hyperbolic_Kmeans

References I

H. Djeddal, L. Touzari, A. Giovanidis, C.-D. Phung, S. Secci
Hyperbolic K-means for traffic-aware clustering in cloud and virtualized RANs.
Elsevier Computer Communications, 176, pp. 258–271. (2021)

A. Pizzinat, P. Chanclou, F. Saliou, and T. Diallo.
Things you should know about fronthaul.
J. Lightwave Technol., 33 (5) pp. 1077–1083. (2015)

L. Chen, D. Yang, D. Zhang, C. Wang, J. Li, T.-M.-T. Nguyen.
Deep mobile traffic forecast and complementary base station clustering for C-RAN optimization.
J. Netw. Comput. Appl., no.121, pp. 59–69. (2018)

W. da Silva Coelho, A. Benhamiche, N. Perrot, S. Secci.
On the impact of novel function mappings, sharing policies, and split settings in network slice design.
16th International Conference on Network and Service Management (CNSM), pp. 1–9. (2020)

References II

D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, M. Boguná.
Hyperbolic geometry of complex networks.
Phys. Rev. E 82 (2010)

M. Nickel and D. Kiela.
Poincaré embeddings for learning hierarchical representations,
Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS?17, , Curran Associates Inc., Red Hook, NY, USA, (2017)

H. Hajri, H. Zaatiti, G. Hébrail, P. Aknin.
Apprentissage automatique sur des données de type graphe utilisant le plongement de Poincaré et les algorithmes stochastiques riemanniens.
Conférence Nationale D'Intelligence Artificielle, (2019)