

Klaszterezés telekommunikációs adatokon

Kolok Balázs Csegő

2020.december.17.

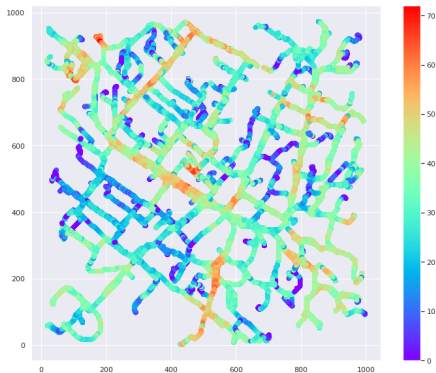
- 5G hálózatok
- Konfigurálható paraméterek száma jelentősen nőtt az előzőgenerációkhoz képest - nehezen kezelhető az eddigi analitikus módszerekkel
- régebbi projekt: A3 esemény előrejelzése

- szimulációs környezet, $1000m \times 1000m$
- tetszőleges számú cellát telepíthetünk
- a felhasználók szabályos időközönként reportolnak
- 600011 darab report
- RSRP lista: 339 hosszú mérési lista
- a mérések 0 – 97-ig terjedő logaritmusos skálán, kevés nem nulla érték
- az első 57 méréssel foglalkoztunk, amelyek egy frekvencia tartományban vannak

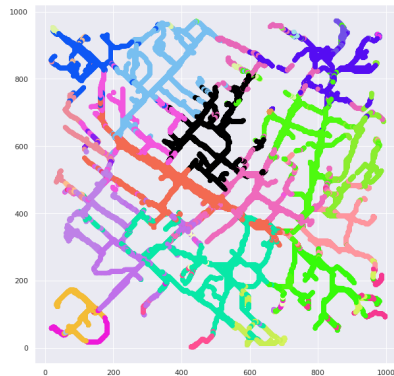
	pos_x	pos_y	UID	Timestamp	rsrp_list
0	537.1	595.5	1351	1601996388439000	[42.0130000000000005, 21.2279999999999994, 10.94...
1	939.0	440.9	1139	1601996388439000	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
2	372.0	465.3	1193	1601996388440000	[19.1809999999999997, 5.6269999999999995, 43.842...

ábra. Az első három felhasználói report

Az adatokról



(a) RSRP listák maximális mérései

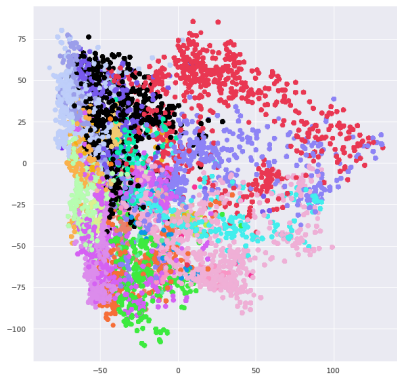


(b) RSRP listák legerősebb cella szerint csoportosítva

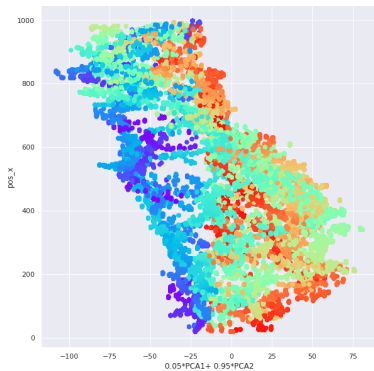
ábra. Szimulációs környezet

- RSRP listák klaszterezése X_1, \dots, X_n klaszterekbe
- felhasználók lokalizálása, hozzárendelése X_1, \dots, X_n klaszterekhez
- trajektóriák alapján a következő klaszter prediktálása

Dimenziócsökkentés: PCA

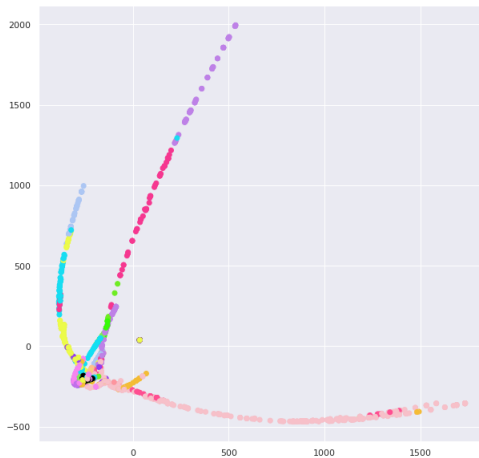


(a) Első és második főkomponens szerinti vetítés



(b) Főkomponensek és a koordináták összefüggése

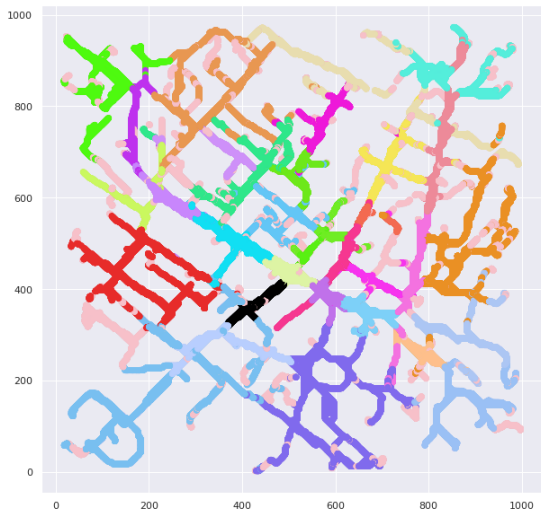
ábra. PCA



ábra. Az adatok 2D-s beágyazása Isomap algoritmussal

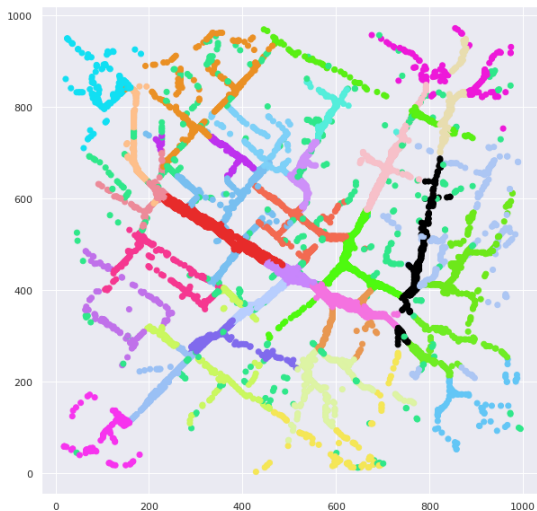
- cél: két cella határán finomabb klaszterezés mint cella közepén
- klaszterezés az egész adathalmazon
- klaszterezés cellákon belül
- kipróbált algoritmusok: K-Means, K-Medoids, Agglomerative Clustering, BIRCH

Klaszterezés az egész adathalmazon

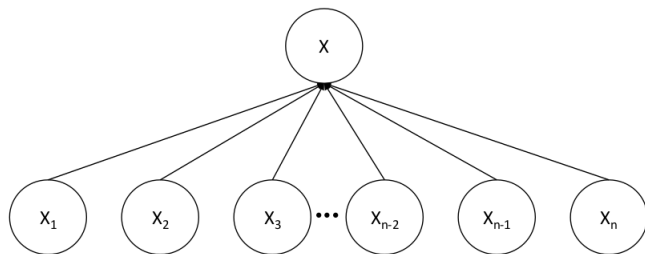


ábra. K-Means, $k = 34$

Klaszterezés az egész adathalmazon



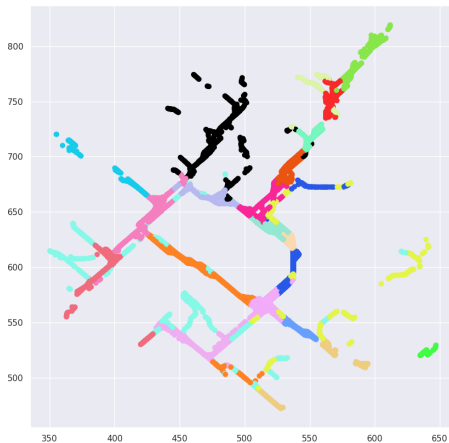
ábra. BIRCH által adott klaszterek, $k = 34$



Klaszterezés az egész térképen

Klaszterezés a cellán belül

Agglomerative Clustering



ábra. AC

	ARI	$maxdist_{av}$	std_{av}
K-means	0.2696	196.55	76.24
K-means PCA után	0.2275	198.97	80.02
K-means Isomap után	0.0755	201.57	85.31
K-medoids	0.2121	200.15	82.64
AggClus	0.2811	171.25	71.94
BIRCH	0.3104	194.67	77.19

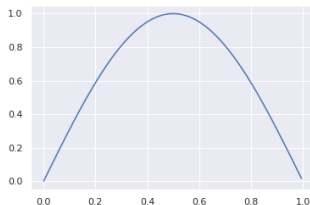
táblázat. A klaszterező módszerek összehasonlítása

Speciális távolságfüggvény

- kívánt klaszterezés eléréséhez valamilyen speciális távolságfüggvényre van szükségünk

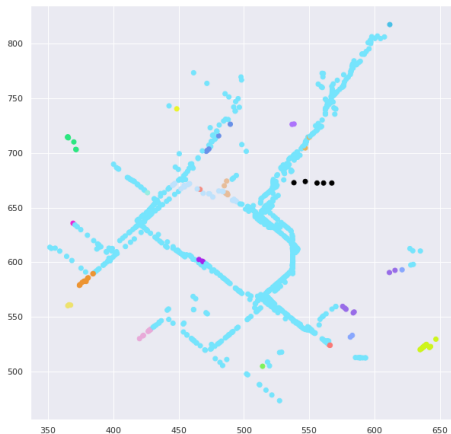
$$d_s(v, u) = \left(\sum_{i=1}^n (w_s(v_i, u_i) |v_i - u_i|)^r \right)^{\frac{1}{r}}$$

$$w_s(a, b) = \sin \frac{(a + b)\pi}{194}$$



ábra. w_s súlyfüggvény

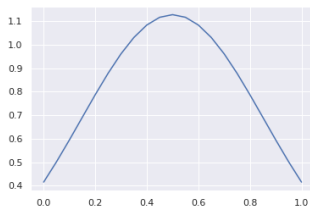
Klaszterezés speciális távolságfüggvénnyel



ábra. w_s súlyfüggvénnyel

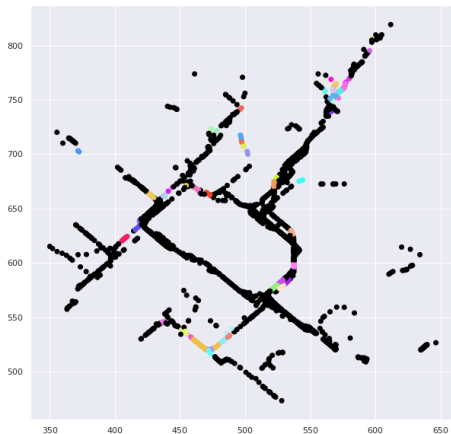
Speciális távolságfüggvény

$$d_{\delta}(v, u) = \left(\sum_{i=1}^n (w_{\delta}(v_i, u_i) |v_i - u_i|)^r \right)^{\frac{1}{r}}$$
$$w_{\delta}(a, b) = \frac{1}{|c|\sqrt{\pi}} \exp \left(- \left(\frac{\frac{a+b}{194} - \frac{1}{2}}{c} \right)^2 \right)$$



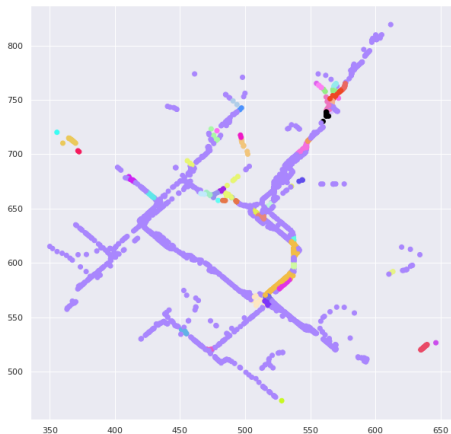
ábra. w_{δ} súlyfüggvény $c = 0.5$

Klaszterezés speciális távolságfüggvénnyel



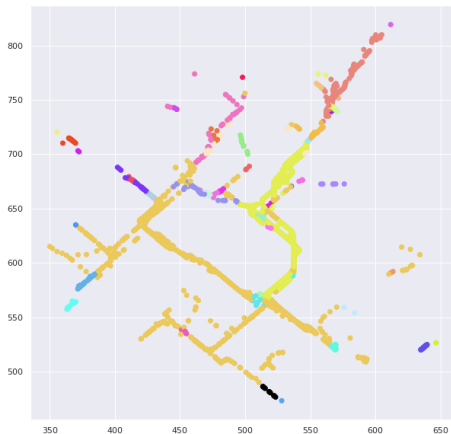
ábra. w_δ súlyfüggvénnyel $c = 0.1$

Klaszterezés speciális távolságfüggvénnyel



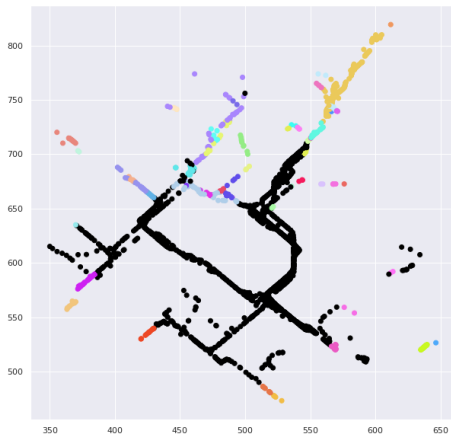
ábra. w_δ súlyfüggvénnyel $c = 0.2$

Klaszterezés speciális távolságfüggvénnyel



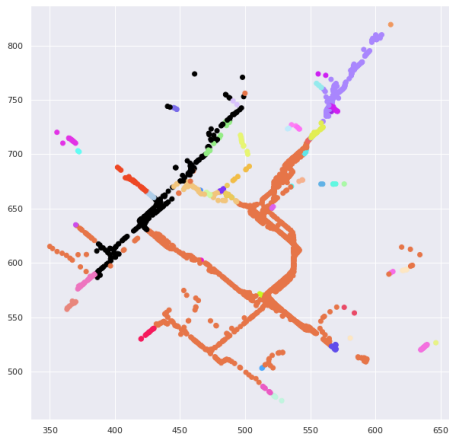
ábra. w_δ súlyfüggvénnyel $c = 0.3$

Klaszterezés speciális távolságfüggvénnyel



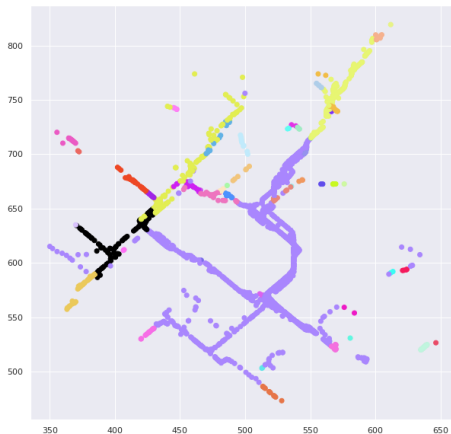
ábra. w_δ súlyfüggvénnyel $c = 0.4$

Klaszterezés speciális távolságfüggvénnyel



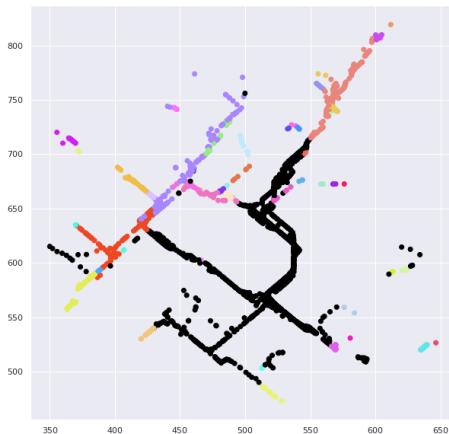
ábra. w_δ súlyfüggvénnyel $c = 0.5$

Klaszterezés speciális távolságfüggvénnyel



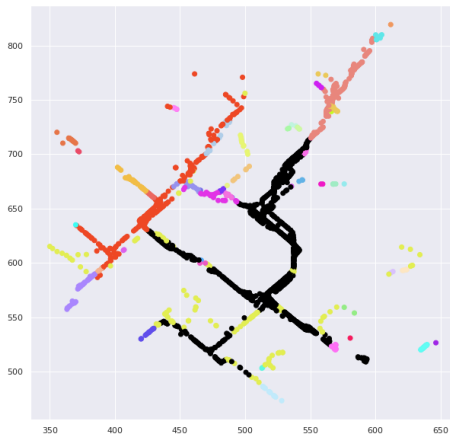
ábra. w_δ súlyfüggvénnyel $c = 0.6$

Klaszterezés speciális távolságfüggvénnyel



ábra. w_δ súlyfüggvénnyel $c = 0.7$

Klaszterezés speciális távolságfüggvénnyel



ábra. w_δ súlyfüggvénnyel $c = 0.8$