

# EXPERIMENTATION

## Effect number of clusters

**Index:** Arxiv

**Rang freq:** 0.1-0.3

**Numwords:** 200

**Max iters:** 20

**Iterations:** 3

**Clusters:** 8

**Comment:** Tenim un conjunt molt petit de clusters per tant convergeix molt fàcilment, per la qual cosa anirem incrementant els clústers per intentar veure alguna tendència, ja que es el cas base.

**Index:** Arxiv

**Rang freq:** 0.1-0.3

**Numwords:** 200

**Max iters:** 20

**Iterations:** 3

**Clusters:** 16

**Comment:** Aconsegüim el mateix resultat que a l'experiment anterior, llavors hem de incrementar el número de clusters per observar resultats

**Index:** Arxiv

**Rang freq:** 0.1-0.3

**Numwords** 200

**Max iters:** 20

**Iterations:** 4

**Clusters:** 128

**Comment:** Finalment, amb aquest algoritme, encara que no ha acabat de convergir pel gran número de clusters, amb aquest numero de iteracions, veiem com en aquesta vegada és a les 4 iteracions quan el nostre algorisme convergeix, per tant hi ha un augment en la complexitat de càlcul.

# Effect vocabulary size

Index: Arxiv

Rang freq: 0.1-0.5

**Numwords** 50

**Max iters:** 20

**Iterations:** 3

**Clusters:** 64

**Time:** 56.63 seconds

**Clusters:** 1

**Numwords** 100

**Max iters:** 20

**Iterations:** 3

**Clusters:** 64

**Time:** 96.21 seconds

**Clusters:** 1

**Numwords** 200

**Max iters:** 20

**Iterations:** 5

**Clusters:** 64

**Time:** 177.26 seconds

**Clusters:** 1

**Numwords** 500

**Max iters:** 20

**Iterations:** 5

**Clusters:** 64

**Time:** 244.1 seconds

**Clusters:** 1

## Effect computation time (#map-reduce)

**Dataset:**

**Index:** arxiv

**Rang freq:** 0.1-0.3

**NWords:** 100

**Clusters:** 128

Iter\nCore	1	2	4	8	16
1	97.44 s	53.16 s	31.49 s	25.24 s	27.54 s
2	320.07 s	177.90 s	105.93 s	84.59 s	86.93 s
3	4.63 s	2.96 s	2.37 s	2.46 s	2.95 s
4					
5					

# Proof of correctness

**Dataset:**

**Index:** arxiv

**NWords:** 1000

**NClust:** 128

**Max\_iter:** 30

**Iter:** 5

**Time:** 973 sec

**Result:** (Process Result)

*Look at prototype0.txt*

=>

*Look at prototype1.txt*

=>

CLASS76

[(0.9234338747099768, 'we'), (0.669953596287703, 'which'), (0.6177494199535963, 'can'), (0.6160092807424594, 'use'), (0.5951276102088167, 'our')]

CLASS54

[(0.9309090909090909, 'we'), (0.8363636363636363, 'neural'), (0.8072727272727273, 'network'), (0.7418181818181818, 'use'), (0.6727272727272727, 'which')]

CLASS65

[(0.9108900842383031, 'we'), (0.6920734801583274, 'which'), (0.5641936466050949, 'from'), (0.5525220744950776, 'can'), (0.5515071551811631, 'result')]

CLASS79

[(0.962877030162413, 'we'), (0.8004640371229699, 'mass'), (0.7935034802784223, 'from'), (0.7935034802784223, 'accret'), (0.691415313225058, 'odot')]

CLASS92

[(0.8909231230973716, 'we'), (0.6927925224027783, 'from'), (0.6897054409810058, 'use'), (0.6156583629893239, 'which'), (0.5604767825751404, 'model')]

CLASS116

[(0.963855421686747, 'we'), (0.7349397590361446, 'show'), (0.7108433734939759, 'our'), (0.6385542168674698, 'which'), (0.6385542168674698, 'can')]

CLASS125

[(1.0, 'adversari'), (0.9487179487179487, 'we'), (0.8974358974358975, 'neural'), (0.8717948717948718, 'network'), (0.7948717948717948, 'can')]

CLASS16

[(0.9605263157894737, 'we'), (0.7763157894736842, 'show'), (0.6973684210526315, 'our'), (0.6842105263157895, 'can'), (0.5921052631578947, 'from')]

CLASS25

[(0.7222222222222222, 'we'), (0.6666666666666666, 'which'), (0.6388888888888888, 'from'), (0.6111111111111112, 'state'), (0.5833333333333334, 'energi')]

CLASS27

[(0.914833215046132, 'we'), (0.6687958362905133, 'which'), (0.6229004021764845, 'from'), (0.5874142417790396, 'use'), (0.5803170096995505, 'can')]

CLASS70

[(0.8395784543325527, 'we'), (0.6830601092896175, 'use'), (0.6693989071038251, 'from'), (0.6026541764246682, 'which'), (0.5497007546187874, 'can')]

CLASS0

[(0.8772324216310728, 'we'), (0.6722867490945422, 'from'), (0.6323217184963157, 'which'), (0.5747470962907456, 'use'), (0.5101785937304858, 'model')]

=>

CLASS76

[(0.8064516129032258, 'we'), (0.6612903225806451, 'our'), (0.6612903225806451, 'model'), (0.6290322580645161, 'use'), (0.6129032258064516, 'show')]

CLASS92

[(0.9137168141592921, 'we'), (0.7074115044247787, 'from'), (0.7050147492625368, 'use'), (0.6071165191740413, 'which'), (0.6050884955752213, 'model')]

CLASS65

[(0.8976266383280198, 'we'), (0.6767876119629573, 'which'), (0.5942513030717069, 'from'), (0.5933404179950408, 'use'), (0.5711755477961642, 'can')]

CLASS27

[(0.8941118605958404, 'we'), (0.6930157391793142, 'from'), (0.6675098369870713, 'use'), (0.6066610455311973, 'which'), (0.5354834176503653, 'can')]

CLASS70

[(0.8522928806688624, 'we'), (0.6960983025082341, 'use'), (0.6755763871294654, 'from'), (0.6078033949835318, 'which'), (0.5600456042563973, 'can')]

CLASS0

[(0.8799332299988076, 'we'), (0.6748539406223918, 'from'), (0.6279957076427805, 'which'), (0.5738643138190056, 'use'), (0.507452009061643, 'model')]

=>

CLASS65

[(0.8891140602582497, 'we'), (0.6540530846484935, 'from'), (0.6349175035868005, 'which'), (0.6347919655667145, 'use'), (0.5451757532281205, 'can')]

=>

CLASS65

[(0.8891140602582497, 'we'), (0.6540530846484935, 'from'), (0.6349175035868005, 'which'), (0.6347919655667145, 'use'), (0.5451757532281205, 'can')]