OT-SA Taxonomy

We will show how to build a taxonomy of clusters from multiple samples. The data set in this example is the Baron Pancreas dataset. It only has 4 samples which allows us to visualize the taxonomy with a dendrogram. The method by which clusters are aligned here is regular OT.

We first load the Mosek Matlab package:

```
addpath C:/Users/sebastian/Mosek/10.2/toolbox/r2017a
```

Load the formatted Baron Pancreas dataset:

```
load baronpc ;
dim = 50; %number of PCs
numPat = length(stride); %number of samples in dataset
celltypes= unique(cellnames); %unique cell types present in dataset
numcells = length(celltypes); %number of unique cell types.
```

We now apply the OT-RMC algorithm for each sample. The output for a sample is a similarity matrix with respect to the clusters present in that sample.

```
lambda2=0;
lambda = .075;
matchSelf = cell(numPat,1);
for i =1:numPat
    nclust=stride(i); %number of clusters in sample i
    start = sum(stride(1:i-1))+1;
    ms = supp(1:dim,start:start+nclust-1); %mean vectors of clusters in sample i
    vars = supp(dim+1:dim^2+dim,start:start+nclust-1); %cov matrices of clusters in sample i
    p = ww(start:start+nclust-1);%proportions of clusters in sample i
    cost = CostMat(ms,ms,vars,vars,nclust,nclust); %compute cost matrix
    cost = real(cost/max(cost,[],"all"));  %standardize
    [~,res]=OT(cost,p,p);
    xx=res.sol.itr.xx;
    gammaij=reshape(xx,[nclust,nclust]);
    %normalization
    gammaijcol =  gammaij./ max(abs(gammaij), [], 1);
    gammaijrow =  gammaij./ max(abs(gammaij), [], 2);
    matchSelf{i} = (gammaijcol+gammaijrow)/2;
end
```

```
MOSEK Version 10.2.5 (Build date: 2024-9-17 14:14:12)
Copyright (c) MOSEK ApS, Denmark WWW: mosek.com
Platform: Windows/64-X86

Problem
  Name                   :
  Objective sense        : minimize
  Type                   : LO (linear optimization problem)
  Constraints            : 28
  Affine conic cons.     : 0
  Disjunctive cons.      : 0
  Cones                  : 0
  Scalar variables       : 196
  Matrix variables       : 0
```

```
  Integer variables      : 0

Optimizer started.
Presolve started.
Linear dependency checker started.
Linear dependency checker terminated.
Eliminator started.
Freed constraints in eliminator : 0
Eliminator terminated.
Eliminator - tries                  : 1                 time                   : 0.00
Lin. dep.  - tries                  : 1                 time                   : 0.00
Lin. dep.  - primal attempts        : 1                 successes              : 1
Lin. dep.  - dual attempts          : 0                 successes              : 0
Lin. dep.  - primal deps.           : 1                 dual deps.             : 0
Presolve terminated. Time: 0.00
Optimizer  - threads                : 4
Optimizer  - solved problem         : the primal
Optimizer  - Constraints            : 27
Optimizer  - Cones                  : 0
Optimizer  - Scalar variables       : 196               conic                  : 0
Optimizer  - Semi-definite variables: 0                 scalarized             : 0
Factor     - setup time             : 0.00
Factor     - dense det. time        : 0.00              GP order time          : 0.00
Factor     - nonzeros before factor : 209               after factor           : 287
Factor     - dense dim.             : 0                 flops                  : 4.68e+03
ITE PFEAS    DFEAS    GFEAS    PRSTATUS   POBJ              DOBJ              MU       TIME
0   1.7e+00  4.4e-01  1.2e+01  0.00e+00   1.121735064e+01   -9.976372414e-01  1.6e-01  0.02
1   5.3e-01  1.4e-01  3.8e+00  1.38e+00   2.572686545e+00   -2.554371438e-01  4.9e-02  0.03
2   7.0e-02  1.8e-02  5.1e-01  1.70e+00   2.263839249e-01   -3.307175336e-02  6.6e-03  0.03
3   8.0e-03  2.1e-03  5.8e-02  1.29e+00   2.127138034e-02   -3.999807663e-03  7.4e-04  0.03
4   2.7e-04  1.7e-04  2.4e-03  1.17e+00   7.332874268e-04   -2.341756479e-04  3.1e-05  0.03
5   8.2e-07  5.0e-07  7.2e-06  1.02e+00   2.413527227e-06   -4.640102693e-07  9.4e-08  0.05
6   8.2e-11  5.0e-11  7.2e-10  1.00e+00   6.848507925e-08   6.819732199e-08   9.4e-12  0.05
Basis identification started.
Primal basis identification phase started.
Primal basis identification phase terminated. Time: 0.00
Dual basis identification phase started.
Dual basis identification phase terminated. Time: 0.00
Basis identification terminated. Time: 0.00
Optimizer terminated. Time: 0.09


Interior-point solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 6.8485079253e-08    nrm: 5e-01    Viol.  con: 3e-11    var: 0e+00
  Dual.    obj: 6.8197321991e-08    nrm: 1e+00    Viol.  con: 0e+00    var: 2e-11

Basic solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 6.8250548272e-08    nrm: 5e-01    Viol.  con: 0e+00    var: 0e+00
  Dual.    obj: 6.8197321992e-08    nrm: 2e+00    Viol.  con: 0e+00    var: 1e-09
Optimizer summary
  Optimizer                 -                       time: 0.09
    Interior-point          - iterations : 6        time: 0.06
      Basis identification  -                       time: 0.00
        Primal              - iterations : 0        time: 0.00
        Dual                - iterations : 13       time: 0.00
        Clean primal        - iterations : 0        time: 0.00
        Clean dual          - iterations : 0        time: 0.00
    Simplex                 -                       time: 0.00
      Primal simplex        - iterations : 0        time: 0.00
      Dual simplex          - iterations : 0        time: 0.00
    Mixed integer           - relaxations: 0        time: 0.00
```

```
MOSEK Version 10.2.5 (Build date: 2024-9-17 14:14:12)
Copyright (c) MOSEK ApS, Denmark WWW: mosek.com
Platform: Windows/64-X86

Problem
  Name                   :
  Objective sense        : minimize
  Type                   : LO (linear optimization problem)
  Constraints            : 28
  Affine conic cons.     : 0
  Disjunctive cons.      : 0
  Cones                  : 0
  Scalar variables       : 196
  Matrix variables       : 0
  Integer variables      : 0

Optimizer started.
Presolve started.
Linear dependency checker started.
Linear dependency checker terminated.
Eliminator started.
Freed constraints in eliminator : 0
Eliminator terminated.
Eliminator - tries                  : 1                 time                   : 0.00
Lin. dep.  - tries                  : 1                 time                   : 0.00
Lin. dep.  - primal attempts        : 1                 successes              : 1
Lin. dep.  - dual attempts          : 0                 successes              : 0
Lin. dep.  - primal deps.           : 1                 dual deps.             : 0
Presolve terminated. Time: 0.00
Optimizer  - threads                : 4
Optimizer  - solved problem         : the primal
Optimizer  - Constraints            : 27
Optimizer  - Cones                  : 0
Optimizer  - Scalar variables       : 196               conic                  : 0
Optimizer  - Semi-definite variables: 0                 scalarized             : 0
Factor     - setup time             : 0.00
Factor     - dense det. time        : 0.00              GP order time          : 0.00
Factor     - nonzeros before factor : 209               after factor           : 287
Factor     - dense dim.             : 0                 flops                  : 4.68e+03
ITE PFEAS    DFEAS    GFEAS    PRSTATUS   POBJ              DOBJ              MU       TIME
0   1.7e+00  3.1e-01  1.2e+01  0.00e+00   1.082275814e+01   -1.024190053e+00  1.8e-01  0.01
1   2.7e-01  4.9e-02  1.9e+00  1.90e+00   9.784529916e-01   -9.439682157e-02  2.8e-02  0.03
2   2.6e-02  4.9e-03  1.9e-01  1.39e+00   7.194526261e-02   -1.194871059e-02  2.8e-03  0.03
3   4.1e-03  7.5e-04  2.9e-02  1.28e+00   9.856645279e-03   -1.404026706e-03  4.2e-04  0.03
4   1.1e-04  5.1e-05  9.0e-04  1.15e+00   2.046802322e-04   -1.179068294e-04  1.3e-05  0.05
5   7.1e-07  3.4e-07  6.0e-06  1.05e+00   1.468701830e-06   -6.511446844e-07  8.9e-08  0.05
6   7.6e-11  3.7e-11  6.5e-10  1.00e+00   5.784279741e-08   5.761609784e-08   9.6e-12  0.05
Basis identification started.
Primal basis identification phase started.
Primal basis identification phase terminated. Time: 0.00
Dual basis identification phase started.
Dual basis identification phase terminated. Time: 0.01
Basis identification terminated. Time: 0.01
Optimizer terminated. Time: 0.09


Interior-point solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 5.7842797413e-08    nrm: 3e-01    Viol.  con: 3e-11    var: 0e+00
  Dual.    obj: 5.7616097839e-08    nrm: 1e+00    Viol.  con: 0e+00    var: 1e-11

Basic solution summary
```

```
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 5.7691787587e-08    nrm: 3e-01    Viol.  con: 0e+00    var: 0e+00
  Dual.    obj: 5.7616097839e-08    nrm: 2e+00    Viol.  con: 0e+00    var: 5e-08
Optimizer summary
  Optimizer                 -                        time: 0.09
    Interior-point          - iterations : 6         time: 0.06
      Basis identification  -                        time: 0.01
        Primal              - iterations : 0         time: 0.00
        Dual                - iterations : 13        time: 0.01
        Clean primal        - iterations : 0         time: 0.00
        Clean dual          - iterations : 0         time: 0.00
    Simplex                 -                        time: 0.00
      Primal simplex        - iterations : 0         time: 0.00
      Dual simplex          - iterations : 0         time: 0.00
    Mixed integer           - relaxations: 0         time: 0.00


MOSEK Version 10.2.5 (Build date: 2024-9-17 14:14:12)
Copyright (c) MOSEK ApS, Denmark WWW: mosek.com
Platform: Windows/64-X86

Problem
  Name                   :
  Objective sense        : minimize
  Type                   : LO (linear optimization problem)
  Constraints            : 28
  Affine conic cons.     : 0
  Disjunctive cons.      : 0
  Cones                  : 0
  Scalar variables       : 196
  Matrix variables       : 0
  Integer variables      : 0

Optimizer started.
Presolve started.
Linear dependency checker started.
Linear dependency checker terminated.
Eliminator started.
Freed constraints in eliminator : 0
Eliminator terminated.
Eliminator - tries                  : 1                 time                   : 0.00
Lin. dep.  - tries                  : 1                 time                   : 0.02
Lin. dep.  - primal attempts        : 1                 successes              : 1
Lin. dep.  - dual attempts          : 0                 successes              : 0
Lin. dep.  - primal deps.           : 1                 dual deps.             : 0
Presolve terminated. Time: 0.02
Optimizer  - threads                : 4
Optimizer  - solved problem         : the primal
Optimizer  - Constraints            : 27
Optimizer  - Cones                  : 0
Optimizer  - Scalar variables       : 196               conic                  : 0
Optimizer  - Semi-definite variables: 0                 scalarized             : 0
Factor     - setup time             : 0.00
Factor     - dense det. time        : 0.00              GP order time          : 0.00
Factor     - nonzeros before factor : 209               after factor           : 287
Factor     - dense dim.             : 0                 flops                  : 4.68e+03
ITE PFEAS    DFEAS    GFEAS    PRSTATUS   POBJ              DOBJ              MU       TIME
0   1.7e+00  5.5e-01  1.5e+01  0.00e+00   1.393184117e+01   -1.068392929e+00  1.6e-01  0.03
1   3.6e-01  1.2e-01  3.3e+00  1.06e+00   2.446021571e+00   -2.250300595e-01  3.6e-02  0.03
2   4.2e-02  1.2e-02  3.8e-01  1.38e+00   2.040778770e-01   -3.582528235e-02  4.1e-03  0.05
3   4.4e-03  9.8e-04  3.9e-02  1.31e+00   1.674574097e-02   -3.644862047e-03  4.2e-04  0.05
4   1.0e-04  6.7e-05  1.2e-03  1.19e+00   3.741952955e-04   -2.050730830e-04  1.3e-05  0.05
5   9.4e-08  6.0e-08  1.1e-06  1.02e+00   3.392677121e-07   -1.760652310e-07  1.2e-08  0.05
6   1.0e-11  6.5e-12  1.2e-10  1.00e+00   1.095654685e-07   1.095093081e-07   1.3e-12  0.05
```

```
Basis identification started.
Primal basis identification phase started.
Primal basis identification phase terminated. Time: 0.00
Dual basis identification phase started.
Dual basis identification phase terminated. Time: 0.01
Basis identification terminated. Time: 0.01
Optimizer terminated. Time: 0.09


Interior-point solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 1.0956546852e-07    nrm: 4e-01    Viol.  con: 5e-12    var: 0e+00
  Dual.    obj: 1.0950930807e-07    nrm: 1e+00    Viol.  con: 0e+00    var: 3e-12

Basic solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 1.0954043279e-07    nrm: 4e-01    Viol.  con: 0e+00    var: 0e+00
  Dual.    obj: 1.0950930801e-07    nrm: 2e+00    Viol.  con: 0e+00    var: 9e-09
Optimizer summary
  Optimizer                 -                        time: 0.09
    Interior-point          - iterations : 6         time: 0.06
      Basis identification  -                        time: 0.01
        Primal              - iterations : 0         time: 0.00
        Dual                - iterations : 13        time: 0.01
        Clean primal        - iterations : 0         time: 0.00
        Clean dual          - iterations : 0         time: 0.00
    Simplex                 -                        time: 0.00
      Primal simplex        - iterations : 0         time: 0.00
      Dual simplex          - iterations : 0         time: 0.00
    Mixed integer           - relaxations: 0         time: 0.00


MOSEK Version 10.2.5 (Build date: 2024-9-17 14:14:12)
Copyright (c) MOSEK ApS, Denmark WWW: mosek.com
Platform: Windows/64-X86

Problem
  Name                   :
  Objective sense        : minimize
  Type                   : LO (linear optimization problem)
  Constraints            : 28
  Affine conic cons.     : 0
  Disjunctive cons.      : 0
  Cones                  : 0
  Scalar variables       : 196
  Matrix variables       : 0
  Integer variables      : 0

Optimizer started.
Presolve started.
Linear dependency checker started.
Linear dependency checker terminated.
Eliminator started.
Freed constraints in eliminator : 0
Eliminator terminated.
Eliminator - tries                  : 1                 time                   : 0.00
Lin. dep.  - tries                  : 1                 time                   : 0.02
Lin. dep.  - primal attempts        : 1                 successes              : 1
Lin. dep.  - dual attempts          : 0                 successes              : 0
Lin. dep.  - primal deps.           : 1                 dual deps.             : 0
Presolve terminated. Time: 0.02
Optimizer  - threads                : 4
Optimizer  - solved problem         : the primal
```

```
Optimizer  - Constraints            : 27
Optimizer  - Cones                  : 0
Optimizer  - Scalar variables       : 196              conic                  : 0
Optimizer  - Semi-definite variables: 0                scalarized             : 0
Factor     - setup time             : 0.01
Factor     - dense det. time        : 0.00             GP order time          : 0.00
Factor     - nonzeros before factor : 209              after factor           : 287
Factor     - dense dim.             : 0                flops                  : 4.68e+03
ITE PFEAS    DFEAS    GFEAS    PRSTATUS   POBJ              DOBJ              MU       TIME
0   1.7e+00  5.1e-01  1.5e+01  0.00e+00   1.334854790e+01   -1.081459940e+00  1.7e-01  0.03
1   3.6e-01  1.1e-01  3.1e+00  1.20e+00   2.201014664e+00   -2.006778581e-01  3.6e-02  0.05
2   4.2e-02  1.3e-02  3.7e-01  1.35e+00   2.036286790e-01   -2.453735077e-02  4.2e-03  0.05
3   4.6e-03  9.3e-04  3.8e-02  1.25e+00   1.715037941e-02   -3.559793348e-03  4.4e-04  0.05
4   2.2e-04  6.5e-05  2.0e-03  1.27e+00   6.248034125e-04   -2.923466449e-04  2.3e-05  0.05
5   4.2e-07  1.2e-07  3.8e-06  1.06e+00   1.351620741e-06   -3.382616587e-07  4.3e-08  0.05
6   4.2e-11  1.2e-11  3.8e-10  1.00e+00   5.917840012e-08   5.900885571e-08   4.3e-12  0.05
Basis identification started.
Primal basis identification phase started.
Primal basis identification phase terminated. Time: 0.00
Dual basis identification phase started.
Dual basis identification phase terminated. Time: 0.00
Basis identification terminated. Time: 0.00
Optimizer terminated. Time: 0.09


Interior-point solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 5.9178400122e-08    nrm: 4e-01    Viol.  con: 2e-11    var: 0e+00
  Dual.    obj: 5.9008855706e-08    nrm: 1e+00    Viol.  con: 0e+00    var: 6e-12

Basic solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 5.9048715346e-08    nrm: 4e-01    Viol.  con: 0e+00    var: 0e+00
  Dual.    obj: 5.9008855656e-08    nrm: 2e+00    Viol.  con: 0e+00    var: 9e-09
Optimizer summary
  Optimizer                 -                        time: 0.09
    Interior-point          - iterations : 6         time: 0.06
      Basis identification  -                        time: 0.00
        Primal              - iterations : 0         time: 0.00
        Dual                - iterations : 13        time: 0.00
        Clean primal        - iterations : 0         time: 0.00
        Clean dual          - iterations : 0         time: 0.00
    Simplex                 -                        time: 0.00
      Primal simplex        - iterations : 0         time: 0.00
      Dual simplex          - iterations : 0         time: 0.00
    Mixed integer           - relaxations: 0         time: 0.00
```

We compute the matching matrices for the different pairwise sample combinations. For each pair of samples we obtain a similarity matrix

```
combs = nchoosek(1:numPat,2);
matchM = cell(length(combs),1);
for i=1:length(combs)
    nclust1=stride(combs(i,1));  %number of clusters in 1st sample
    nclust2=stride(combs(i,2));  %number of clusters in 2nd sample
    start1 = sum(stride(1:combs(i,1)-1))+1;   %how many columns to ignore +1
    start2 = sum(stride(1:combs(i,2)-1))+1;
    ms1 = supp(1:dim,start1:start1+nclust1-1); %mean vectors of clusters in first sample
    ms2 = supp(1:dim,start2:start2+nclust2-1);
```

```
    vars1=supp(dim+1:dim+dim^2,start1:start1+nclust1-1); %cov matrices of clusters in first sam
    vars2 = supp(dim+1:dim+dim^2,start2:start2+nclust2-1);
    p1 = ww(start1:start1+nclust1-1);  %proportions of clusters in sample 1
    p2 = ww(start2:start2+nclust2-1);
    cost = CostMat(ms1,ms2,vars1,vars2,nclust1,nclust2); %cost matrix computation
    cost = real(cost/max(cost,[],"all"));
    [~,res]=OT(cost,p1,p2);
    xx=res.sol.itr.xx;
    gammaij=reshape(xx,[nclust1,nclust2]);
    %normalization
    if sum(sum(gammaij)) ~=0
        gammaijcol =  gammaij./ max(abs(gammaij), [], 1);
        gammaijrow =  gammaij./ max(abs(gammaij), [], 2);
        matchM{i} = (gammaijcol+gammaijrow)/2;
    else
        matchM{i} = gammaij;
    end
end
```

```
MOSEK Version 10.2.5 (Build date: 2024-9-17 14:14:12)
Copyright (c) MOSEK ApS, Denmark WWW: mosek.com
Platform: Windows/64-X86

Problem
  Name                   :
  Objective sense        : minimize
  Type                   : LO (linear optimization problem)
  Constraints            : 28
  Affine conic cons.     : 0
  Disjunctive cons.      : 0
  Cones                  : 0
  Scalar variables       : 196
  Matrix variables       : 0
  Integer variables      : 0

Optimizer started.
Presolve started.
Linear dependency checker started.
Linear dependency checker terminated.
Eliminator started.
Freed constraints in eliminator : 0
Eliminator terminated.
Eliminator - tries                  : 1                 time                   : 0.00
Lin. dep.  - tries                  : 1                 time                   : 0.00
Lin. dep.  - primal attempts        : 1                 successes              : 1
Lin. dep.  - dual attempts          : 0                 successes              : 0
Lin. dep.  - primal deps.           : 1                 dual deps.             : 0
Presolve terminated. Time: 0.00
Optimizer  - threads                : 4
Optimizer  - solved problem         : the primal
Optimizer  - Constraints            : 27
Optimizer  - Cones                  : 0
Optimizer  - Scalar variables       : 196               conic                  : 0
Optimizer  - Semi-definite variables: 0                 scalarized             : 0
Factor     - setup time             : 0.00
Factor     - dense det. time        : 0.00              GP order time          : 0.00
Factor     - nonzeros before factor : 209               after factor           : 287
Factor     - dense dim.             : 0                 flops                  : 4.68e+03
ITE PFEAS    DFEAS    GFEAS    PRSTATUS   POBJ              DOBJ              MU       TIME
0   1.9e+00  1.2e-01  1.4e+01  0.00e+00   1.268799652e+01   -6.826861994e-01  1.7e-01  0.01
1   4.7e-01  2.9e-02  3.4e+00  3.19e+00   1.597515029e+00   2.300236060e-01   4.1e-02  0.01
```

```
2   2.2e-01  1.4e-02  1.6e+00  1.42e+00   8.822250278e-01   3.294305934e-01   1.9e-02  0.03
3   3.9e-02  2.4e-03  2.8e-01  1.46e+00   4.543950590e-01   3.771362880e-01   3.4e-03  0.03
4   8.6e-03  5.3e-04  6.1e-02  1.26e+00   3.992315161e-01   3.844139444e-01   7.5e-04  0.03
5   1.8e-03  1.1e-04  1.3e-02  1.17e+00   3.889030037e-01   3.860807292e-01   1.6e-04  0.03
6   4.4e-04  2.7e-05  3.1e-03  1.12e+00   3.870214420e-01   3.863671288e-01   3.8e-05  0.03
7   2.5e-05  1.5e-06  1.8e-04  1.02e+00   3.864625466e-01   3.864258400e-01   2.2e-06  0.03
8   7.1e-07  4.4e-08  5.1e-06  1.00e+00   3.864301402e-01   3.864290796e-01   6.2e-08  0.03
9   7.6e-11  4.7e-12  5.4e-10  1.00e+00   3.864291918e-01   3.864291917e-01   6.6e-12  0.03
Basis identification started.
Primal basis identification phase started.
Primal basis identification phase terminated. Time: 0.00
Dual basis identification phase started.
Dual basis identification phase terminated. Time: 0.00
Basis identification terminated. Time: 0.02
Optimizer terminated. Time: 0.08


Interior-point solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 3.8642919181e-01    nrm: 5e-01    Viol.  con: 2e-11    var: 0e+00
  Dual.    obj: 3.8642919170e-01    nrm: 1e+00    Viol.  con: 0e+00    var: 1e-12

Basic solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 3.8642919171e-01    nrm: 5e-01    Viol.  con: 6e-17    var: 0e+00
  Dual.    obj: 3.8642919170e-01    nrm: 1e+00    Viol.  con: 0e+00    var: 1e-08
Optimizer summary
  Optimizer                 -                        time: 0.08
    Interior-point          - iterations : 9         time: 0.05
      Basis identification  -                        time: 0.02
        Primal              - iterations : 0         time: 0.00
        Dual                - iterations : 0         time: 0.00
        Clean primal        - iterations : 0         time: 0.00
        Clean dual          - iterations : 0         time: 0.00
    Simplex                 -                        time: 0.00
      Primal simplex        - iterations : 0         time: 0.00
      Dual simplex          - iterations : 0         time: 0.00
    Mixed integer           - relaxations: 0         time: 0.00


MOSEK Version 10.2.5 (Build date: 2024-9-17 14:14:12)
Copyright (c) MOSEK ApS, Denmark WWW: mosek.com
Platform: Windows/64-X86

Problem
  Name                   :
  Objective sense        : minimize
  Type                   : LO (linear optimization problem)
  Constraints            : 28
  Affine conic cons.     : 0
  Disjunctive cons.      : 0
  Cones                  : 0
  Scalar variables       : 196
  Matrix variables       : 0
  Integer variables      : 0

Optimizer started.
Presolve started.
Linear dependency checker started.
Linear dependency checker terminated.
Eliminator started.
Freed constraints in eliminator : 0
Eliminator terminated.
```

```
Eliminator - tries                : 1                 time                   : 0.00
Lin. dep.  - tries                : 1                 time                   : 0.00
Lin. dep.  - primal attempts      : 1                 successes              : 1
Lin. dep.  - dual attempts        : 0                 successes              : 0
Lin. dep.  - primal deps.         : 1                 dual deps.             : 0
Presolve terminated. Time: 0.01
Optimizer  - threads              : 4
Optimizer  - solved problem       : the primal
Optimizer  - Constraints          : 27
Optimizer  - Cones                : 0
Optimizer  - Scalar variables     : 196               conic                  : 0
Optimizer  - Semi-definite variables: 0               scalarized             : 0
Factor     - setup time           : 0.00
Factor     - dense det. time      : 0.00              GP order time          : 0.00
Factor     - nonzeros before factor : 209             after factor           : 287
Factor     - dense dim.           : 0                 flops                  : 4.68e+03
ITE PFEAS    DFEAS    GFEAS    PRSTATUS   POBJ              DOBJ              MU       TIME
0   1.9e+00  1.1e-01  1.5e+01  0.00e+00   1.372254016e+01   -7.123448206e-01  1.7e-01  0.01
1   4.9e-01  2.7e-02  3.8e+00  3.56e+00   1.584299799e+00   2.196694186e-01   4.4e-02  0.05
2   1.4e-01  7.6e-03  1.0e+00  1.46e+00   6.363877889e-01   3.284988441e-01   1.2e-02  0.05
3   2.6e-02  1.5e-03  2.0e-01  1.31e+00   4.014970253e-01   3.519709868e-01   2.3e-03  0.05
4   7.2e-03  4.0e-04  5.5e-02  1.21e+00   3.708099789e-01   3.584259379e-01   6.4e-04  0.05
5   1.2e-03  6.5e-05  8.9e-03  1.15e+00   3.617234232e-01   3.598671502e-01   1.0e-04  0.05
6   5.9e-05  3.3e-06  4.6e-04  1.02e+00   3.602170245e-01   3.601213500e-01   5.3e-06  0.06
7   5.8e-07  3.3e-08  4.5e-06  1.00e+00   3.601357940e-01   3.601348550e-01   5.2e-08  0.06
8   1.4e-09  7.8e-11  1.1e-08  1.00e+00   3.601350240e-01   3.601350217e-01   1.2e-10  0.06
9   2.0e-13  7.8e-15  1.1e-12  1.00e+00   3.601350219e-01   3.601350219e-01   1.2e-14  0.08
Basis identification started.
Primal basis identification phase started.
Primal basis identification phase terminated. Time: 0.06
Dual basis identification phase started.
Dual basis identification phase terminated. Time: 0.02
Basis identification terminated. Time: 0.08
Optimizer terminated. Time: 0.22


Interior-point solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 3.6013502187e-01    nrm: 5e-01    Viol.  con: 4e-14    var: 0e+00
  Dual.    obj: 3.6013502187e-01    nrm: 8e-01    Viol.  con: 0e+00    var: 2e-15

Basic solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 3.6013502187e-01    nrm: 5e-01    Viol.  con: 1e-16    var: 0e+00
  Dual.    obj: 3.6013502187e-01    nrm: 8e-01    Viol.  con: 0e+00    var: 7e-12
Optimizer summary
  Optimizer                 -                        time: 0.22
    Interior-point          - iterations : 9         time: 0.17
      Basis identification  -                        time: 0.08
        Primal              - iterations : 0         time: 0.06
        Dual                - iterations : 0         time: 0.02
        Clean primal        - iterations : 0         time: 0.00
        Clean dual          - iterations : 0         time: 0.00
    Simplex                 -                        time: 0.00
      Primal simplex        - iterations : 0         time: 0.00
      Dual simplex          - iterations : 0         time: 0.00
    Mixed integer           - relaxations: 0         time: 0.00


MOSEK Version 10.2.5 (Build date: 2024-9-17 14:14:12)
Copyright (c) MOSEK ApS, Denmark WWW: mosek.com
Platform: Windows/64-X86
```

```
Problem
  Name                   :
  Objective sense        : minimize
  Type                   : LO (linear optimization problem)
  Constraints            : 28
  Affine conic cons.     : 0
  Disjunctive cons.      : 0
  Cones                  : 0
  Scalar variables       : 196
  Matrix variables       : 0
  Integer variables      : 0

Optimizer started.
Presolve started.
Linear dependency checker started.
Linear dependency checker terminated.
Eliminator started.
Freed constraints in eliminator : 0
Eliminator terminated.
Eliminator - tries                  : 1                 time                   : 0.00
Lin. dep.  - tries                  : 1                 time                   : 0.00
Lin. dep.  - primal attempts        : 1                 successes              : 1
Lin. dep.  - dual attempts          : 0                 successes              : 0
Lin. dep.  - primal deps.           : 1                 dual deps.             : 0
Presolve terminated. Time: 0.02
Optimizer  - threads                : 4
Optimizer  - solved problem         : the primal
Optimizer  - Constraints            : 27
Optimizer  - Cones                  : 0
Optimizer  - Scalar variables       : 196               conic                  : 0
Optimizer  - Semi-definite variables: 0                 scalarized             : 0
Factor     - setup time             : 0.00
Factor     - dense det. time        : 0.00              GP order time          : 0.00
Factor     - nonzeros before factor : 209               after factor           : 287
Factor     - dense dim.             : 0                 flops                  : 4.68e+03
ITE PFEAS    DFEAS    GFEAS    PRSTATUS   POBJ              DOBJ              MU       TIME
0   1.9e+00  1.2e-01  1.5e+01  0.00e+00   1.411364457e+01   -6.926314022e-01  1.7e-01  0.03
1   5.4e-01  3.5e-02  4.3e+00  3.72e+00   1.730735547e+00   2.257353780e-01   4.8e-02  0.05
2   1.6e-01  1.0e-02  1.3e+00  1.38e+00   6.954790948e-01   3.074386989e-01   1.4e-02  0.05
3   2.7e-02  1.7e-03  2.1e-01  1.28e+00   3.866737163e-01   3.310979283e-01   2.4e-03  0.05
4   7.4e-03  4.8e-04  5.8e-02  1.32e+00   3.487017049e-01   3.355688522e-01   6.6e-04  0.05
5   6.7e-04  4.4e-05  5.3e-03  1.22e+00   3.377742004e-01   3.367150213e-01   6.1e-05  0.06
6   3.5e-05  2.3e-06  2.7e-04  1.01e+00   3.368527677e-01   3.367981082e-01   3.1e-06  0.06
7   3.4e-07  2.2e-08  2.7e-06  1.00e+00   3.368014725e-01   3.368009361e-01   3.0e-08  0.06
8   3.4e-11  2.2e-12  2.7e-10  1.00e+00   3.368009909e-01   3.368009908e-01   3.0e-12  0.06
Basis identification started.
Primal basis identification phase started.
Primal basis identification phase terminated. Time: 0.00
Dual basis identification phase started.
Dual basis identification phase terminated. Time: 0.01
Basis identification terminated. Time: 0.01
Optimizer terminated. Time: 0.11


Interior-point solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 3.3680099087e-01    nrm: 5e-01    Viol.  con: 7e-12    var: 0e+00
  Dual.    obj: 3.3680099082e-01    nrm: 7e-01    Viol.  con: 0e+00    var: 4e-13

Basic solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 3.3680099082e-01    nrm: 5e-01    Viol.  con: 7e-18    var: 0e+00
  Dual.    obj: 3.3680099082e-01    nrm: 7e-01    Viol.  con: 0e+00    var: 5e-09
```

```
Optimizer summary
  Optimizer                 -                        time: 0.11
    Interior-point          - iterations : 8         time: 0.08
      Basis identification  -                        time: 0.01
        Primal              - iterations : 0         time: 0.00
        Dual                - iterations : 0         time: 0.01
        Clean primal        - iterations : 0         time: 0.00
        Clean dual          - iterations : 0         time: 0.00
    Simplex                 -                        time: 0.00
      Primal simplex        - iterations : 0         time: 0.00
      Dual simplex          - iterations : 0         time: 0.00
    Mixed integer           - relaxations: 0         time: 0.00


MOSEK Version 10.2.5 (Build date: 2024-9-17 14:14:12)
Copyright (c) MOSEK ApS, Denmark WWW: mosek.com
Platform: Windows/64-X86

Problem
  Name                   :
  Objective sense        : minimize
  Type                   : LO (linear optimization problem)
  Constraints            : 28
  Affine conic cons.     : 0
  Disjunctive cons.      : 0
  Cones                  : 0
  Scalar variables       : 196
  Matrix variables       : 0
  Integer variables      : 0

Optimizer started.
Presolve started.
Linear dependency checker started.
Linear dependency checker terminated.
Eliminator started.
Freed constraints in eliminator : 0
Eliminator terminated.
Eliminator - tries                  : 1                 time                   : 0.00
Lin. dep.  - tries                  : 1                 time                   : 0.00
Lin. dep.  - primal attempts        : 1                 successes              : 1
Lin. dep.  - dual attempts          : 0                 successes              : 0
Lin. dep.  - primal deps.           : 1                 dual deps.             : 0
Presolve terminated. Time: 0.00
Optimizer  - threads                : 4
Optimizer  - solved problem         : the primal
Optimizer  - Constraints            : 27
Optimizer  - Cones                  : 0
Optimizer  - Scalar variables       : 196               conic                  : 0
Optimizer  - Semi-definite variables: 0                 scalarized             : 0
Factor     - setup time             : 0.00
Factor     - dense det. time        : 0.00              GP order time          : 0.00
Factor     - nonzeros before factor : 209               after factor           : 287
Factor     - dense dim.             : 0                 flops                  : 4.68e+03
ITE PFEAS    DFEAS    GFEAS    PRSTATUS   POBJ              DOBJ              MU       TIME
0   1.9e+00  2.1e-01  1.5e+01  0.00e+00   1.415848620e+01   -6.143545484e-01  1.8e-01  0.01
1   5.7e-01  6.4e-02  4.5e+00  4.29e+00   1.746813893e+00   3.270480327e-01   5.4e-02  0.01
2   1.2e-01  1.4e-02  9.9e-01  1.33e+00   6.434993241e-01   3.772571921e-01   1.2e-02  0.03
3   3.1e-02  3.5e-03  2.4e-01  1.36e+00   4.479671525e-01   3.925924472e-01   2.9e-03  0.03
4   5.2e-03  5.9e-04  4.1e-02  1.24e+00   4.052927317e-01   3.970696975e-01   4.9e-04  0.03
5   1.4e-03  1.5e-04  1.1e-02  1.18e+00   3.996443735e-01   3.976536024e-01   1.3e-04  0.03
6   2.9e-04  3.3e-05  2.3e-03  1.07e+00   3.982030979e-01   3.977913377e-01   2.8e-05  0.03
7   1.2e-05  1.4e-06  9.8e-05  1.01e+00   3.978323154e-01   3.978144766e-01   1.2e-06  0.03
8   5.2e-09  5.9e-10  4.1e-08  1.00e+00   3.978154487e-01   3.978154413e-01   4.9e-10  0.03
Basis identification started.
Primal basis identification phase started.
```

```
Primal basis identification phase terminated. Time: 0.00
Dual basis identification phase started.
Dual basis identification phase terminated. Time: 0.02
Basis identification terminated. Time: 0.02
Optimizer terminated. Time: 0.08


Interior-point solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 3.9781544869e-01    nrm: 4e-01    Viol.  con: 9e-10    var: 0e+00
  Dual.    obj: 3.9781544126e-01    nrm: 7e-01    Viol.  con: 0e+00    var: 1e-10

Basic solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 3.9781544184e-01    nrm: 4e-01    Viol.  con: 1e-16    var: 0e+00
  Dual.    obj: 3.9781544126e-01    nrm: 7e-01    Viol.  con: 0e+00    var: 3e-07
Optimizer summary
  Optimizer                 -                        time: 0.08
    Interior-point          - iterations : 8         time: 0.05
      Basis identification  -                        time: 0.02
        Primal              - iterations : 0         time: 0.00
        Dual                - iterations : 0         time: 0.02
        Clean primal        - iterations : 0         time: 0.00
        Clean dual          - iterations : 0         time: 0.00
    Simplex                 -                        time: 0.00
      Primal simplex        - iterations : 0         time: 0.00
      Dual simplex          - iterations : 0         time: 0.00
    Mixed integer           - relaxations: 0         time: 0.00


MOSEK Version 10.2.5 (Build date: 2024-9-17 14:14:12)
Copyright (c) MOSEK ApS, Denmark WWW: mosek.com
Platform: Windows/64-X86

Problem
  Name                   :
  Objective sense        : minimize
  Type                   : LO (linear optimization problem)
  Constraints            : 28
  Affine conic cons.     : 0
  Disjunctive cons.      : 0
  Cones                  : 0
  Scalar variables       : 196
  Matrix variables       : 0
  Integer variables      : 0

Optimizer started.
Presolve started.
Linear dependency checker started.
Linear dependency checker terminated.
Eliminator started.
Freed constraints in eliminator : 0
Eliminator terminated.
Eliminator - tries                  : 1                 time                   : 0.00
Lin. dep.  - tries                  : 1                 time                   : 0.00
Lin. dep.  - primal attempts        : 1                 successes              : 1
Lin. dep.  - dual attempts          : 0                 successes              : 0
Lin. dep.  - primal deps.           : 1                 dual deps.             : 0
Presolve terminated. Time: 0.02
Optimizer  - threads                : 4
Optimizer  - solved problem         : the primal
Optimizer  - Constraints            : 27
Optimizer  - Cones                  : 0
```

```
Optimizer  - Scalar variables       : 196              conic                  : 0
Optimizer  - Semi-definite variables: 0                scalarized             : 0
Factor     - setup time             : 0.00
Factor     - dense det. time        : 0.00             GP order time          : 0.00
Factor     - nonzeros before factor : 209              after factor           : 287
Factor     - dense dim.             : 0                flops                  : 4.68e+03
ITE PFEAS    DFEAS    GFEAS    PRSTATUS   POBJ              DOBJ              MU       TIME
0   1.9e+00  2.0e-01  1.5e+01  0.00e+00   1.448100040e+01   -6.346398643e-01  1.8e-01  0.03
1   5.2e-01  5.4e-02  4.2e+00  4.16e+00   1.679141086e+00   3.460975762e-01   4.9e-02  0.05
2   1.3e-01  1.3e-02  1.0e+00  1.28e+00   6.777261578e-01   3.916089298e-01   1.2e-02  0.05
3   2.8e-02  3.0e-03  2.3e-01  1.30e+00   4.653370254e-01   4.092696915e-01   2.7e-03  0.05
4   8.6e-03  9.0e-04  7.0e-02  1.23e+00   4.280328336e-01   4.128461875e-01   8.2e-04  0.05
5   9.2e-04  9.6e-05  7.4e-03  1.25e+00   4.154070819e-01   4.140497261e-01   8.7e-05  0.05
6   1.4e-04  1.4e-05  1.1e-03  1.08e+00   4.143548861e-01   4.141540396e-01   1.3e-05  0.05
7   1.1e-05  1.1e-06  8.5e-05  1.03e+00   4.141916311e-01   4.141761446e-01   1.0e-06  0.05
8   1.0e-07  1.1e-08  8.5e-07  1.00e+00   4.141769098e-01   4.141767562e-01   1.0e-08  0.05
9   1.1e-11  1.1e-12  8.6e-11  1.00e+00   4.141767637e-01   4.141767637e-01   1.0e-12  0.05
Basis identification started.
Primal basis identification phase started.
Primal basis identification phase terminated. Time: 0.02
Dual basis identification phase started.
Dual basis identification phase terminated. Time: 0.00
Basis identification terminated. Time: 0.02
Optimizer terminated. Time: 0.08


Interior-point solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 4.1417676367e-01    nrm: 4e-01    Viol.  con: 2e-12    var: 0e+00
  Dual.    obj: 4.1417676365e-01    nrm: 8e-01    Viol.  con: 0e+00    var: 2e-13

Basic solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 4.1417676365e-01    nrm: 4e-01    Viol.  con: 1e-17    var: 0e+00
  Dual.    obj: 4.1417676365e-01    nrm: 8e-01    Viol.  con: 0e+00    var: 8e-10
Optimizer summary
  Optimizer                 -                        time: 0.08
    Interior-point          - iterations : 9         time: 0.06
      Basis identification  -                        time: 0.02
        Primal              - iterations : 0         time: 0.02
        Dual                - iterations : 0         time: 0.00
        Clean primal        - iterations : 0         time: 0.00
        Clean dual          - iterations : 0         time: 0.00
    Simplex                 -                        time: 0.00
      Primal simplex        - iterations : 0         time: 0.00
      Dual simplex          - iterations : 0         time: 0.00
    Mixed integer           - relaxations: 0         time: 0.00


MOSEK Version 10.2.5 (Build date: 2024-9-17 14:14:12)
Copyright (c) MOSEK ApS, Denmark WWW: mosek.com
Platform: Windows/64-X86

Problem
  Name                   :
  Objective sense        : minimize
  Type                   : LO (linear optimization problem)
  Constraints            : 28
  Affine conic cons.     : 0
  Disjunctive cons.      : 0
  Cones                  : 0
  Scalar variables       : 196
  Matrix variables       : 0
```

```
  Integer variables      : 0

Optimizer started.
Presolve started.
Linear dependency checker started.
Linear dependency checker terminated.
Eliminator started.
Freed constraints in eliminator : 0
Eliminator terminated.
Eliminator - tries                  : 1                 time                   : 0.00
Lin. dep.  - tries                  : 1                 time                   : 0.00
Lin. dep.  - primal attempts        : 1                 successes              : 1
Lin. dep.  - dual attempts          : 0                 successes              : 0
Lin. dep.  - primal deps.           : 1                 dual deps.             : 0
Presolve terminated. Time: 0.00
Optimizer  - threads                : 4
Optimizer  - solved problem         : the primal
Optimizer  - Constraints            : 27
Optimizer  - Cones                  : 0
Optimizer  - Scalar variables       : 196               conic                  : 0
Optimizer  - Semi-definite variables: 0                 scalarized             : 0
Factor     - setup time             : 0.00
Factor     - dense det. time        : 0.00              GP order time          : 0.00
Factor     - nonzeros before factor : 209               after factor           : 287
Factor     - dense dim.             : 0                 flops                  : 4.68e+03
ITE PFEAS    DFEAS    GFEAS    PRSTATUS   POBJ              DOBJ              MU       TIME
0   1.9e+00  1.1e-01  1.7e+01  0.00e+00   1.606574097e+01   -7.739208945e-01  1.8e-01  0.02
1   3.0e-01  1.7e-02  2.7e+00  3.20e+00   1.323177771e+00   2.349941742e-01   2.8e-02  0.03
2   8.3e-02  4.8e-03  7.6e-01  1.36e+00   5.585328347e-01   3.071749189e-01   7.8e-03  0.03
3   3.1e-02  1.8e-03  2.8e-01  1.14e+00   4.137462696e-01   3.222389300e-01   3.0e-03  0.03
4   5.0e-03  2.9e-04  4.5e-02  1.20e+00   3.411821789e-01   3.286411097e-01   4.7e-04  0.03
5   6.5e-04  3.8e-05  6.0e-03  1.26e+00   3.317081676e-01   3.302648941e-01   6.2e-05  0.03
6   6.5e-05  3.7e-06  5.9e-04  1.04e+00   3.305020670e-01   3.303593410e-01   6.1e-06  0.03
7   1.0e-06  5.8e-08  9.2e-06  1.01e+00   3.303856739e-01   3.303834664e-01   9.5e-08  0.03
8   1.0e-10  6.0e-12  9.5e-10  1.00e+00   3.303837308e-01   3.303837306e-01   9.9e-12  0.03
Basis identification started.
Primal basis identification phase started.
Primal basis identification phase terminated. Time: 0.00
Dual basis identification phase started.
Dual basis identification phase terminated. Time: 0.00
Basis identification terminated. Time: 0.00
Optimizer terminated. Time: 0.05


Interior-point solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 3.3038373083e-01    nrm: 4e-01    Viol.  con: 3e-11    var: 0e+00
  Dual.    obj: 3.3038373060e-01    nrm: 1e+00    Viol.  con: 0e+00    var: 1e-12

Basic solution summary
  Problem status  : PRIMAL_AND_DUAL_FEASIBLE
  Solution status : OPTIMAL
  Primal.  obj: 3.3038373062e-01    nrm: 4e-01    Viol.  con: 1e-16    var: 0e+00
  Dual.    obj: 3.3038373060e-01    nrm: 1e+00    Viol.  con: 0e+00    var: 1e-08
Optimizer summary
  Optimizer                 -                       time: 0.05
    Interior-point          - iterations : 8        time: 0.03
      Basis identification  -                       time: 0.00
        Primal              - iterations : 0        time: 0.00
        Dual                - iterations : 0        time: 0.00
        Clean primal        - iterations : 0        time: 0.00
        Clean dual          - iterations : 0        time: 0.00
    Simplex                 -                       time: 0.00
      Primal simplex        - iterations : 0        time: 0.00
```

```
  Dual simplex          - iterations : 0         time: 0.00
Mixed integer           - relaxations: 0         time: 0.00
```

We now combine all of the different matching matrices into one big matching matrix, $\boldsymbol{B}$. We refer to $\boldsymbol{B}$ as the overall cluster similarity matrix.

```
Biggammaij = blkdiag(matchSelf{:}); % this functions takes a finite  list of matrices
                                    % and creates a larger matrix witht the
                                    % matrices as diagonal blocks. {:}
                                    % spits out all the matrices from the
                                    % cell array
startPat=1;
endPat = numPat-1;
for j =1:(numPat-1)
    nclust1=stride(j);  %number of clusters for individual j
    start1 = sum(stride(1:j-1))+1;   %how many columns to ignore +1
    endClust=start1+nclust1-1;
    Biggammaij(start1:endClust,endClust+1:end)=Biggammaij(start1:endClust,endClust+1:end)+ [mat
    Biggammaij(endClust+1:end,start1:endClust)=Biggammaij(endClust+1:end,start1:endClust)+ [mat
    startPat = endPat+1;
    endPat = startPat+numPat-2-j;
end
%We bound below very small values in the large matrix:
Biggammaij(abs(Biggammaij) < min(ww)* 1E-7) = min(ww)* 1E-7;
```

Construct Tree from Overall Similarity Matrix

We now need to transform $\mathbf{B}$ into a cluster distance matrix $\mathbf{A}$ from which we construct a taxonomy.

```
A = real(-log(Biggammaij));
A= A/ max(max(A));
B = squareform(A);
tree=linkage(B,'ward');
cellClusters=cluster(tree,"MaxClust", numcells);
```

cellClusters is a vector with a length equivalent to that of the number of clusters. The entries in it give the metacluster label for all of the clusters.

Measure Performance

We compute the ARI, cell-level accuracy, as well as cluster level accuracy.

```
failcount = 0;
fail = 0;
for i=1:numcells
    PLindices = find(cellClusters==i); %indices for clusters with metacluster label i
    predictedLabels = string(cellnames(PLindices)); % Ground truth labels for clusters in metac
    predLabel=mode(categorical(predictedLabels)); %Predicted label for metacluster i
    failcount = failcount+ sum(predictedLabels~=string(predLabel)); %add number of clusters in
    propLabels = ww(PLindices); %cluster proportions
    cumstride = cumsum(stride);
    for j=1:length(PLindices)
```

```
        if predictedLabels(j)~=string(predLabel)
            cellsp=numperclus(min(find(cumstride>= PLindices(j)))); %this is the
            %total number of cells in the sample where cell type j is from.
            fail=fail+round(propLabels(j)*cellsp);
        end
    end

end
%ARI
[RI, ARI] = randindex(string(cellClusters),string(cellnames));
ARI
```

```
ARI = 0.7637
```

```
%CLuster accuracy
1-failcount/length(cellnames)
```

```
ans = 0.8750
```

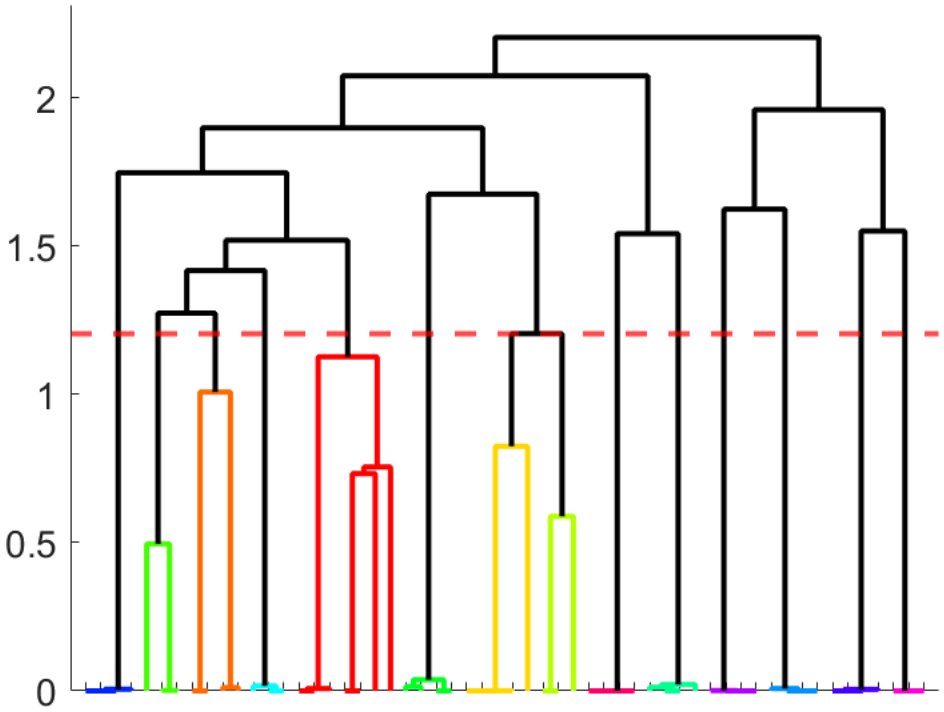
```
%Cell accuracy. %numperclus gives the # of cells per cluster.
1-double(fail)/double(sum(numperclus))
```

```
ans = 0.9940
```

Dendrogram

We can visualize our dendrogram:

```
cutoff = max(tree(end-numcells+2, 3)); % Determine the cutoff height
figure;
H = dendrogram(tree, 0,'ColorThreshold', cutoff);
set(H, 'LineWidth', 2);
set(gca, 'FontSize', 14)
yline(cutoff, 'r--', 'LineWidth', 2); % Add the horizontal line
set(gca, 'XTickLabel', []);
```


2
1.5
1
0.5
0