Untitled

MAT3

2024-10-11

Quarto

Clusters	Shared reactions	Shared MBBs	Clusters	Shared reactions	Shared MBBs
Cluster 1	525	377	Cluster 1	525	377
Cluster 2	243	203	Cluster 2	241	201
Cluster 3	1019	631	Cluster 3	1019	631
Cluster 4	2	2	Cluster 4	2	2

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.5
## v forcats   1.0.0      v stringr   1.5.1
## v ggplot2   3.5.1      v tibble    3.2.1
## v lubridate 1.9.3      v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
dd=read_delim("../data/0a845f74-826e-3b46-aed9-e7ecf74db262/results.csv",delim=",")
```

```
## Rows: 1132 Columns: 3998
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1930): organism, Categories, Groups, mDAG Id, Full Name, R00005(3.5.1.5...
## dbl (2068): R00009(1.11.1.6), R00031(1.10.3.1), R00032(1.13.11.63), R00036(4...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
clusters=unique(dd$Groups)[c(1,2,4,5)]
clusters
```

```
## [1] "MSA Cluster 3|MUN Cluster 3" "MSA Cluster 2|MUN Cluster 2"
## [3] "Cluster 1"                   "Cluster 4"
```

---

[0]MSA-similarity

```
dd=dd%>% filter(Groups %in% clusters)
  reacciones=dd%>% select(-(1:5))
  dim(reacciones)
```

[1] 706 3993

```
  aa=is.na(reacciones)
  Sums=colSums(aa)
  table(Sums==0)
```

```
## 
## FALSE   TRUE
##  3962     31
```

```
  bolas=na.omit(unique(unlist(reacciones)))
  length(bolas)
```

[1] 4034