



vlado

Clustering / World 2023

Github Bavla OpenAlex [<https://github.com/bavla/OpenAlex/blob/main/Countries/pics/world.md>]

C:/Users/vlado/work/OpenAlex/API/world.R

Clustering of the Balassa index matrix

```
> wdir <- "C:/Users/vlado/work/OpenAlex/API"
> setwd(wdir)
>
> library(httr)
> library(jsonlite)
> source("https://raw.githubusercontent.com/bavla/Rnet/master/R/Pajek.R")
> source("OpenAlex4.R")
> library(gplots)
>
> CorEu <- function(W,p=1){
+   D <- W; diag(D) <- 0; n = nrow(D)
+   for(u in 1:(n-1)) for(v in (u+1):n) D[v,u] <- D[u,v] <-
+     sqrt(sum((W[u,]-W[v,])**2) -
+       (W[u,u]-W[v,u])**2 - (W[u,v]-W[v,v])**2 +
+       p*(W[u,u]-W[v,v])**2 + (W[u,v]-W[v,u])**2))
+   return(D)
+ }
> load("./years/Matrix2023.Rdata")
> P <- M; diag(P) <- 0
> Y <- as.character(2023); i <- 1
> D <- rowSums(P); T <- sum(D); n <- nrow(P)
> for(u in 1:(n-1)) for(v in (u+1):n) P[u,v] <- P[v,u] <- P[u,v]*T/D[u]/D[v]
> X <- Z <- log2(P)
> Z[Z == -Inf] <- 0; Z[is.nan(Z)] <- 0
> X[X == -Inf] <- NA; X[is.nan(X)] <- 0
> X[,D==0] <- NA; X[D==0,] <- NA
> t <- hclust(as.dist(CorEu(Z)),method="ward.D")
> plot(t,hang=-1,main="World 2023 / CorrEuclid / Ward",cex=0.3)
> pdf(file=paste("WorldBalassa",Y[i],".pdf",sep=""),width=30,height=30)
> heatmap.2(X,Rowv=as.dendrogram(t),Colv="Rowv",dendrogram="column",
+   scale="none",revC=TRUE,col=bluered(100),na.color="yellow",
+   trace="none",density.info="none",keysize = 0.8,
+   main=paste("World ",Y[i]," / Balassa / Ward",sep=""))
> dev.off()
```

Picture of the clustering of the Balassa index matrix [<https://github.com/bavla/OpenAlex/blob/main/Countries/pics/WorldBalassa2023.pdf>].

Picture of the corresponding intensities of co-authorship [<https://github.com/bavla/OpenAlex/blob/main/Countries/pics/WorldCoA2023.pdf>].

Clustering of coauthorship intensities

```
> P <- M; diag(P) <- 0
> X <- Z <- log2(P)
> Z[Z == -Inf] <- 0; Z[is.nan(Z)] <- 0
> X[X == -Inf] <- NA; X[is.nan(X)] <- 0
> X[,D==0] <- NA; X[D==0,] <- NA
> h <- hclust(as.dist(CorEu(Z)),method="ward.D")
> myPalette <- colorRampPalette(c("white","black"))(n=100)
> pdf(file=paste("WorldCoA",Y[i],".pdf",sep=""),width=30,height=30)
> heatmap.2(X,Rowv=as.dendrogram(t),Colv="Rowv",dendrogram="column",
+   scale="none",revC=TRUE,col=myPalette,na.color="yellow",
+   trace="none",density.info="none",keysize = 0.8,
+   main=paste("World ",Y[i]," / log2 / Ward",sep=""))
> dev.off()
> pdf(file=paste("WorldCoW",Y[i],".pdf",sep=""),width=30,height=30)
> heatmap.2(X,Rowv=as.dendrogram(h),Colv="Rowv",dendrogram="column",
+   scale="none",revC=TRUE,col=myPalette,na.color="yellow",
+   trace="none",density.info="none",keysize = 0.8,
+   main=paste("World ",Y[i]," / log2 w / Ward",sep=""))
> dev.off()
```

Picture of the clustering of intensities [<https://github.com/bavla/OpenAlex/blob/main/Countries/pics/WorldCoW2023.pdf>].

Reordering clusters

A better picture can be obtained by flipping some subtrees in the dendrogram (see Reorder [<https://github.com/bavla/NormNet/blob/main/data/natalija/reorder.md>]).

```
> flip <- function(k,T) {t <- T[k,1]; T[k,1] <- T[k,2]; T[k,2] <- t; return(T)}
> n <- nrow(M); nm <- n-1
> cbind(240:nm,h$merge[240:nm,])
      [,1] [,2] [,3]
[1,] 240 192 220
[2,] 241 229 230
[3,] 242 231 239
[4,] 243 204 240
```

```
[5,] 244 235 241
[6,] 245 242 243
[7,] 246 232 244
[8,] 247 237 238
[9,] 248 233 245
[10,] 249 246 247
[11,] 250 248 249
> s <- h
> h$merge <- flip(186,flip(237,flip(247,h$merge)))
> pdf(file=paste("WorldCoT",Y[i],".pdf",sep=""),width=30,height=30)
> heatmap.2(X,Rowv=as.dendrogram(h),Colv="Rowv",dendrogram="column",
+ scale="none",revC=TRUE,col=myPalette,na.color="yellow",
+ trace="none",density.info="none",keysize = 0.8,
+ main=paste("World ",Y[i]," / log2 w / Ward",sep=""))
> dev.off()
> cbind(235:nm,h$merge[235:nm,])
      [,1] [,2] [,3]
[1,] 235 178 183
[2,] 236 218 227
[3,] 237 186 19
[4,] 238 222 234
[5,] 239 216 236
[6,] 240 192 220
[7,] 241 229 230
[8,] 242 231 239
[9,] 243 204 240
[10,] 244 235 241
[11,] 245 242 243
[12,] 246 232 244
[13,] 247 238 237
[14,] 248 233 245
[15,] 249 246 247
[16,] 250 248 249
```

Picture of the reordered intensity matrix [https://github.com/bavla/OpenAlex/blob/main/Countries/pics/WorldCoT2023.pdf].

Reordering Balassa

```
> toFather <- function(tm){
+   n <- nrow(tm); T <- rep(0,2*n+1)
+   for(i in 1:n){
+     for(j in 1:2){
+       p <- tm[i,j]
+       if(p<0) T[-p] <- i+n+1 else T[n+1+p] <- i+n+1
+     }
+   }
+   return(T)
+ }
>
> minCl <- function(u,v,T){
+   if(min(u,v)==0) return(T[max(u,v)])
+   # cat(u," ",v,":",T[u]," ",T[v],"\n")
+   if(u==v) return(u)
+   return( if(T[u]<T[v]) minCl(T[u],v,T) else minCl(u,T[v],T) )
+ }
>
> F <- toFather(t$merge)
> N <- rownames(M)
> which(N=="SI")
[1] 205
> minCl(which(N=="AN"),which(N=="FR"),F) - n
[1] 239
> minCl(which(N=="IM"),which(N=="FR"),F) - n
[1] 230
> minCl(which(N=="IM"),which(N=="AW"),F) - n
[1] 220
> minCl(which(N=="ST"),which(N=="FR"),F) - n
[1] 225
> minCl(which(N=="TV"),which(N=="FK"),F) - n
[1] 148
> minCl(which(N=="PN"),which(N=="FK"),F) - n
[1] 129
> minCl(which(N=="VG"),which(N=="FK"),F) - n
[1] 78
> minCl(which(N=="SH"),which(N=="FK"),F) - n
[1] 72

> t$merge <- flip(225,flip(220,flip(230,flip(239,flip(250,t$merge)))))
> t$merge <- flip(24,flip(72,flip(78,flip(129,flip(148,t$merge)))))
> pdf(file=paste("WorldBalassaF",Y[i],".pdf",sep=""),width=30,height=30)
> heatmap.2(X,Rowv=as.dendrogram(t),Colv="Rowv",dendrogram="column",
+ scale="none",revC=TRUE,col=bluered(100),na.color="yellow",
+ trace="none",density.info="none",keysize = 0.8,
+ main=paste("World ",Y[i]," / Balassa / Ward",sep=""))
> dev.off()
```

Picture of the reordered Balassa matrix [https://github.com/bavla/OpenAlex/blob/main/Countries/pics/WorldBalassaF2023.pdf].

Europe