



vlado

## Collaboration between countries

### ISO Country Codes

OpenAlex uses ISO\_3166\_alpha-2 Country Codes. ISO [<https://www.iso.org/obp/ui/#search/code/>], WP [[https://en.wikipedia.org/wiki/ISO\\_3166-1\\_alpha-2](https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2)].

I saved the data from [GitHub / georgzoeller \[https://github.com/georgzoeller/iso-3166-1-alpha-2/blob/master/index.js\]](https://github.com/georgzoeller/iso-3166-1-alpha-2/blob/master/index.js) in the file `ISO2codes.csv`.

```
> CD <- read.csv2("ISO2codes.csv", head=TRUE, na.strings="--")
> CN <- CD$name; CC <- CD$code
```

A problem with ISO Country Codes is that they don't contain some ex-countries (YU, SU, ...).

[https://api.openalex.org/works?filter=authorships.institutions.country\\_code:YU](https://api.openalex.org/works?filter=authorships.institutions.country_code:YU) [[https://api.openalex.org/works?filter=authorships.institutions.country\\_code:YU](https://api.openalex.org/works?filter=authorships.institutions.country_code:YU)]

### All countries in OpenAlex

To get the list of all countries used in OpenAlex I tried with

<https://api.openalex.org/works?group-by=authorships.countries> [<https://api.openalex.org/works?group-by=authorships.countries>]

```
> S <- GET("https://api.openalex.org/works?group-by=authorships.countries")
> C <- fromJSON(rawToChar(S$content))$group_by
> head(C)
      key                                     key_display_name  count
1 https://openalex.org/countries/US          United States of America 29098598
2 https://openalex.org/countries/CN              China 13974424
3 https://openalex.org/countries/GB United Kingdom of Great Britain and Northern Ireland 7660952
4 https://openalex.org/countries/DE              Germany 6748267
5 https://openalex.org/countries/JP              Japan 5452295
6 https://openalex.org/countries/FR              France 5233671
> getISO <- function(URLid) substring(URLid, 32)
> dim(C)
[1] 200  3
> WC <- unname(sapply(C$key, getISO))
> WV <- C$count
> setdiff(WC, CC)
character(0)
```

It seems that in this case OpenAlex exports data for only up to 200 most active countries.

### Collaboration matrix

(First version run "Thu May 9 05:10:55 2024")

To get a selected country collaboration list we use the query

<https://api.openalex.org/works?filter=authorships.countries:SI&group-by=authorships.countries> [<https://api.openalex.org/works?filter=authorships.countries:SI&group-by=authorships.countries>]

We can arrange the data in the collaboration matrix `M` with  $M[c1, c2]$  = # of co-authored works by at least one author from country `c1` and country `c2`.

```
> n <- length(CC)
> M <- matrix(0, nrow=n, ncol=n); rownames(M) <- colnames(M) <- CC
> T <- rep(0, n); names(T) <- CC; G <- rep(0, n); names(G) <- CC
> date()
[1] "Fri May 10 03:38:21 2024"
> for(cy in CC){
+   Q <- paste0("https://api.openalex.org/works?filter=",
+     "authorships.countries:", cy,
+     "# ", "publication_year:2010-2020&group-by=authorships.countries")
+     "&group-by=authorships.countries")
+   S <- GET(Q)
+   # S$status_code
+   C <- fromJSON(rawToChar(S$content))
+   D <- C$group_by; T[cy] <- C$meta$count; G[cy] <- C$meta$groups_count
+   J <- unname(sapply(D$key, getISO))
+   if(length(J) > 0){
+     V <- D$count; names(V) <- J
+     for(j in J) M[cy, j] <- V[j]
+   }
+ }
> date()
[1] "Fri May 10 03:39:34 2024"
> save(T, G, M, file="Matrix.Rdata")
```

### Collaboration matrix completion

I expect to obtain data about less active countries by looking at their activity list and using the fact that the matrix is symmetric.

```
> k <- 0; a <- 0
> for(i in 2:n){
+   for(j in 1:(i-1)){
```

```

+   if(M[i,j]!=M[j,i]) { d <- M[i,j]-M[j,i]
+   if(min(M[i,j],M[j,i])==0) { m <- max(M[i,j],M[j,i])
+   M[i,j] <- M[j,i] <- m; k <- k+1
+   } else {a <- a+1; cat(i,j,CC[i],CC[j],M[i,j],M[j,i],d,"\n")}
+   }
+ }
+ }
83 14 DE AU 110756 110755 1
83 61 DE DK 75887 75886 1
110 14 IT AU 57736 57735 1
110 61 IT DK 40142 40141 1
112 22 JP BE 21233 21234 -1
...
238 22 US BE 150982 150988 -6
...
238 83 US DE 702639 702650 -11
...
238 120 US KR 248860 248855 5
...
238 168 US NO 88147 88152 -5
...
238 218 US CH 250499 250494 5
...
240 238 UY US 7042 7043 -1
243 238 VE US 9551 9550 1
246 238 VI US 1046 1045 1
> cat("k =",k," a =",a,"\n")
k = 1636 a = 89

```

I symmetrized asymmetric pairs with one zero-entry by setting it to the other value. It turned out that in the computed matrix, 89 slightly departed from symmetry.

We save the completed collaboration matrix as a Pajek network.

```

> source("https://raw.githubusercontent.com/bavla/Rnet/master/R/Pajek.R")
> matrix2net(M,Net="WorldCo.net")

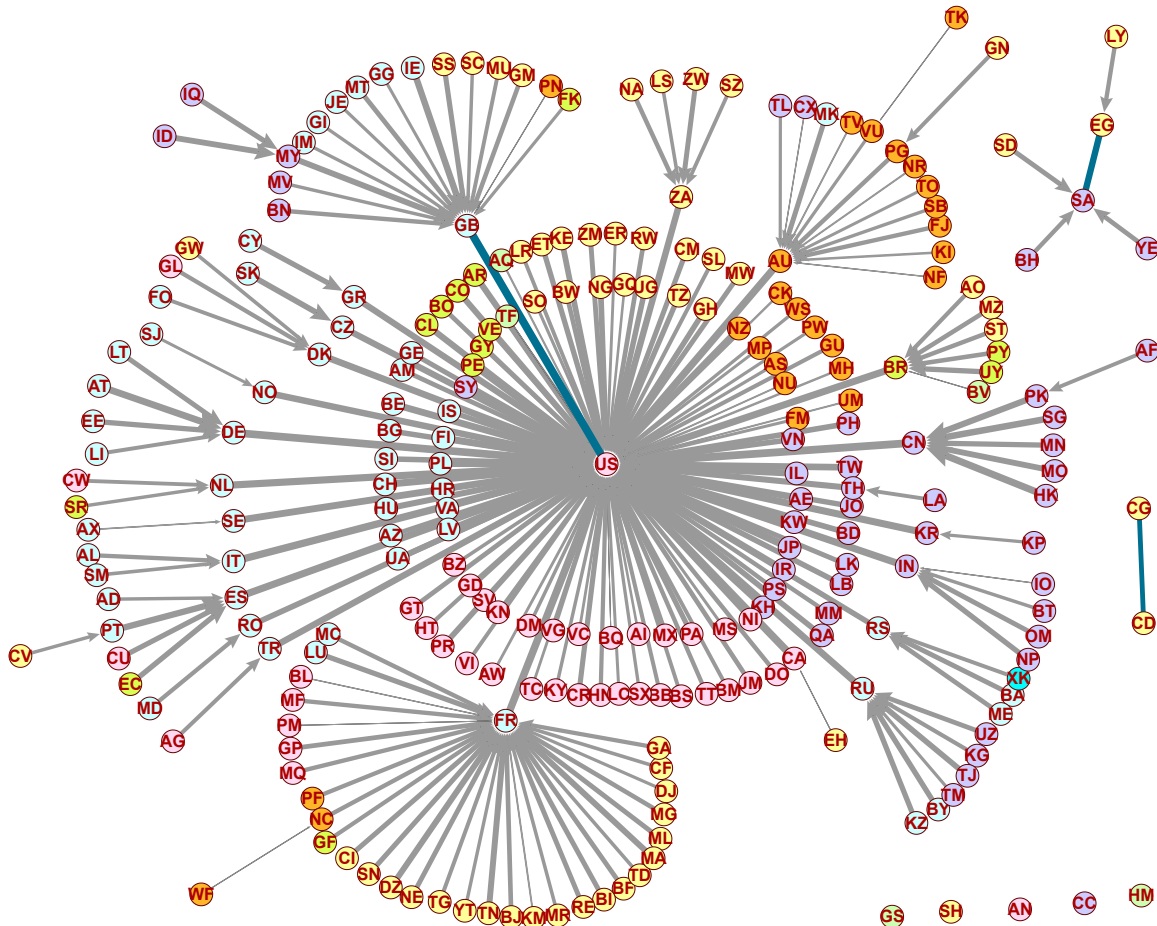
```

A continent partition of countries can be useful. I download the file `country-and-continent-codes-list-csv.csv` [https://gist.github.com/stevewithington/20a69c0b6d2ff846ea5d35e5fc47f26c], created the continent partition and exported it as Pajek file.

```

> H <- read.csv2("country-and-continent-codes-list-csv.csv",sep=";",na.strings="--")
> X <- H$Continent_Code; names(X) <- H$Two_Letter_Country_Code
> Y <- X[CC]
> vecnom2clu(Y,clu="continents.clu")

```



[https://api.openalex.org/works?filter=authorships.institutions.country\\_code:SI](https://api.openalex.org/works?filter=authorships.institutions.country_code:SI) [[https://api.openalex.org/works?filter=authorships.institutions.country\\_code:SI](https://api.openalex.org/works?filter=authorships.institutions.country_code:SI)] | IT

<https://openknowledge.community/tracking-global-access-the-move-to-openalex-and-inclusion-of-2022-data/> [<https://openknowledge.community/tracking-global-access-the-move-to-openalex-and-inclusion-of-2022-data/>]

<https://unstats.un.org/unsd/methodology/m49/> [<https://unstats.un.org/unsd/methodology/m49/>] <https://docs.openalex.org/api-entities/geo/continents> [<https://docs.openalex.org/api-entities/geo/continents>] [https://en.wikipedia.org/wiki/ISO\\_3166-1\\_alpha-2](https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2) [[https://en.wikipedia.org/wiki/ISO\\_3166-1\\_alpha-2](https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2)]

<https://cran.r-project.org/web/packages/countrycode/countrycode.pdf> [<https://cran.r-project.org/web/packages/countrycode/countrycode.pdf>]

---

work/bib/alex/cola.txt · Last modified: 2024/05/11 04:03 by vlado