vlado

Creating a collection

We will work mainly in R and Pajek. We start in R as follows:

```
> library(httr)
> library(jsonlite)
> # source("https://raw.githubusercontent.com/bavla/OpenAlex/main/code/OpenAlex2.R")
> wdir <- "C:/Users/vlado/work/OpenAlex/API"
> setwd(wdir)
> source("OpenAlex2.R")
```

To select a bibliography from OpenAlex we have to construct an appropriate query Q.

Query

In preparation. Skip this section.

```
https://api.openalex.org/works?search=SEARCH&filter=FILTER&select=SELECT
```

```
> wd <- GET("https://api.openalex.org/works",
+   query = list(search=SEARCH,filter=FILTER,select=SELECT))
```

Some elements in the query list can be omitted. Some additional parameters can be used such as per_page, page, sort, etc.

```
> selPub  <- "id,primary_location,title,publication_year,publication_date"
> selRef  <- "biblio,type,language,referenced_works_count,referenced_works"
> selAut  <- "authorships,keywords,countries_distinct_count,cited_by_count"
> selCite <- paste(selPub,selRef,sep=",")
> selAll  <- paste(selCite,selAut,sep=",")
> selInfo <- "id,title,type,publication_year"
```

```
https://api.openalex.org/works?filter=countries_distinct_count:>1&select=id,title
```

```
{
  meta: {
    count: 15907135,
    db_response_time_ms: 90,
    page: 1,
    per_page: 25,
    groups_count: null
  },
  results: [
    {
      id: "https://openalex.org/W1979290264",
      title: "Using thematic analysis in psychology"
    },
    {
      id: "https://openalex.org/W2064675550",
      title: "Long Short-Term Memory"
    },
    {
      ...
    }
  ],
  group_by: [ ]
}
```

IMFM [https://openalex.org/works/I4210106342]

```
https://api.openalex.org/works?filter=authorships.institutions.lineage:i4210106342&select=id,title
https://api.openalex.org/works?filter=institutions.id:I4210106342&select=id,title
https://api.openalex.org/works?filter=institutions.id:I4210106342,publication_year:2010-2020&select=id,title
```

OpenAlex / boolean-searches [https://docs.openalex.org/how-to-use-the-api/get-lists-of-entities/search-entities#boolean-searches]

```
https://api.openalex.org/works?search=(SNA OR "social network" OR "complex network")&select=id,title
```

```
https://api.openalex.org/works?filter=authorships.institutions.country_code:SI,publication_year:2010-2020&group-by=authorships.countries
https://api.openalex.org/works?filter=authorships.institutions.country_code:SI,publication_year:2010-2020,countries_distinct_count:1&select=id,title
https://api.openalex.org/works?filter=publication_year:2010-2020,countries_distinct_count:1&group-by=authorships.countries
https://api.openalex.org/works?filter=authorships.institutions.continent:europe&group-by=authorships.institutions.country_code
```

Saturation

The following query gets from OpenAlex all works co-authored by Vladimir Batagelj [https://openalex.org/authors/A5001676164]

```
> Q <- list(
+   filter="author.id:A5001676164",
+   select=selCite,
+   per_page="200"
+ )
```

To obtain a bibliography of our choice we have to change parameters search and/or filter. Parameters select and page are fixed for OpenAlex2.

This is the saturation part of the construction of a collection of bibliographic networks describing our bibliography. After each run of function `OpenAlex2PajekCite` we read the last version of the citation network into Pajek and apply macro `expNodes` to it. It produces a vector of expansion nodes. Using the vector-Info button in Pajek we get a list of works with the largest input degree. We select an appropriate threshold and extract (select and copy) the upper part of the table into TextPad. In TextPad, we remove other columns and save the list of works as a CSV file (see an example). Using the function `joinLists` we combine the old list of works with the new one and save it for the next step of the saturation procedure.

```
> OpenAlex2PajekCite(Q,1,name="cite",step=100)
OpenAlex2Pajek / Cite - Start Tue May  7 21:36:01 2024
*** OpenAlex2Pajek / Cite - Process Tue May  7 21:36:01 2024
Tue May  7 21:36:03 2024  n = 100
Tue May  7 21:36:03 2024  n = 200
*** OpenAlex2Pajek / Cite - Data Collected Tue May  7 21:36:04 2024
hits: 251 works: 1283
>>> Citation Cite
*** OpenAlex2Pajek / Cite - Stop Tue May  7 21:36:04 2024
> # cite1Ci.net: Pajek + TextPad / deg > 1 -> VB1.csv
> OpenAlex2PajekCite(Q,2,name="cite",listF="VB1.csv",step=100)
OpenAlex2Pajek / Cite - Start Tue May  7 21:53:31 2024
*** OpenAlex2Pajek / Cite - Process Tue May  7 21:53:31 2024
Tue May  7 21:53:33 2024  n = 100
Tue May  7 21:53:33 2024  n = 200
Tue May  7 21:53:53 2024  n = 300
Tue May  7 21:54:25 2024  n = 400
Tue May  7 21:54:56 2024  n = 500
*** OpenAlex2Pajek / Cite - Data Collected Tue May  7 21:55:28 2024
hits: 594 works: 5906
>>> Citation Cite
*** OpenAlex2Pajek / Cite - Stop Tue May  7 21:55:31 2024
> # cite2Ci.net: Pajek + TextPad / deg > 1 -> VB2.csv
> joinLists("VB1.csv","VB2.csv","VB2u.csv")
OnN = 108  O\N = 235  N\O = 935  OUN = 1278
> OpenAlex2PajekCite(Q,3,name="cite",listF="VB2u.csv",step=100)
OpenAlex2Pajek / Cite - Start Tue May  7 22:07:18 2024
*** OpenAlex2Pajek / Cite - Process Tue May  7 22:07:18 2024
Tue May  7 22:07:19 2024  n = 100
Tue May  7 22:07:20 2024  n = 200
Tue May  7 22:07:32 2024  n = 300
...
Tue May  7 22:13:22 2024  n = 1400
Tue May  7 22:13:55 2024  n = 1500
*** OpenAlex2Pajek / Cite - Data Collected Tue May  7 22:14:06 2024
hits: 1529 works: 15076
>>> Citation Cite
*** OpenAlex2Pajek / Cite - Stop Tue May  7 22:14:16 2024
> # cite3Ci.net: Pajek + TextPad / deg>2 -> VB3.csv
> joinLists("VB2u.csv","VB3.csv","VB3u.csv")
OnN = 196  O\N = 1082  N\O = 1073  OUN = 2351
> OpenAlex2PajekCite(Q,4,name="cite",listF="VB3u.csv",step=100)
OpenAlex2Pajek / Cite - Start Tue May  7 22:24:15 2024
*** OpenAlex2Pajek / Cite - Process Tue May  7 22:24:15 2024
Tue May  7 22:24:17 2024  n = 100
Tue May  7 22:24:17 2024  n = 200
Tue May  7 22:24:30 2024  n = 300
'''
Tue May  7 22:34:57 2024  n = 2500
Tue May  7 22:35:29 2024  n = 2600
*** OpenAlex2Pajek / Cite - Data Collected Tue May  7 22:35:30 2024
hits: 2602 works: 24564
>>> Citation Cite
*** OpenAlex2Pajek / Cite - Stop Tue May  7 22:35:59 2024
> # cite4Ci.net: Pajek + TextPad / deg>4 -> VB4.csv
> joinLists("VB3u.csv","VB4.csv","VB4u.csv")
OnN = 292  O\N = 2059  N\O = 270  OUN = 2621
> OpenAlex2PajekCite(Q,5,name="cite",listF="VB4u.csv",step=500)
OpenAlex2Pajek / Cite - Start Tue May  7 22:48:09 2024
*** OpenAlex2Pajek / Cite - Process Tue May  7 22:48:09 2024
Tue May  7 22:49:15 2024  n = 500
Tue May  7 22:51:19 2024  n = 1000
Tue May  7 22:53:31 2024  n = 1500
```

```
Tue May  7 22:55:36 2024  n = 2000
Tue May  7 22:57:35 2024  n = 2500
*** OpenAlex2Pajek / Cite - Data Collected Tue May  7 23:00:34 2024
hits: 2872 works: 26735
>>> Citation Cite
*** OpenAlex2Pajek / Cite - Stop Tue May  7 23:01:12 2024
> # cite5Ci.net: Pajek + TextPad / deg>4 -> VB5.csv
> joinLists("VB4u.csv","VB5.csv","VB5u.csv")
OnN = 377  O\N = 2244  N\O = 126  OUN = 2747
> OpenAlex2PajekCite(Q,6,name="cite",listF="VB5u.csv",step=500)
OpenAlex2Pajek / Cite - Start Tue May  7 23:27:00 2024
*** OpenAlex2Pajek / Cite - Process Tue May  7 23:27:00 2024
Tue May  7 23:28:35 2024  n = 500
Tue May  7 23:31:18 2024  n = 1000
Tue May  7 23:33:52 2024  n = 1500
Tue May  7 23:36:31 2024  n = 2000
Tue May  7 23:40:15 2024  n = 2500
*** OpenAlex2Pajek / Cite - Data Collected Tue May  7 23:43:29 2024
hits: 2998 works: 27629
>>> Citation Cite
*** OpenAlex2Pajek / Cite - Stop Tue May  7 23:44:06 2024
> # cite6Ci.net: Pajek + TextPad / deg>4 -> VB6.csv
> joinLists("VB5u.csv","VB6.csv","VB6u.csv")
OnN = 418  O\N = 2329  N\O = 62  OUN = 2809
> OpenAlex2PajekCite(Q,7,name="cite",listF="VB6u.csv",step=500)
OpenAlex2Pajek / Cite - Start Tue May  7 23:55:55 2024
*** OpenAlex2Pajek / Cite - Process Tue May  7 23:55:55 2024
Tue May  7 23:56:49 2024  n = 500
Tue May  7 23:58:34 2024  n = 1000
Wed May  8 00:00:22 2024  n = 1500
Wed May  8 00:02:16 2024  n = 2000
Wed May  8 00:04:08 2024  n = 2500
Wed May  8 00:06:28 2024  n = 3000
*** OpenAlex2Pajek / Cite - Data Collected Wed May  8 00:06:47 2024
hits: 3060 works: 28097
>>> Citation Cite
*** OpenAlex2Pajek / Cite - Stop Wed May  8 00:07:38 2024
> # cite7Ci.net: Pajek + TextPad / deg>4 -> VB7.csv
> joinLists("VB6u.csv","VB7.csv","VB7u.csv")
OnN = 436  O\N = 2373  N\O = 26  OUN = 2835
> OpenAlex2PajekCite(Q,8,name="cite",listF="VB7u.csv",prop=TRUE,step=500)
OpenAlex2Pajek / Cite - Start Wed May  8 00:41:37 2024
*** OpenAlex2Pajek / Cite - Process Wed May  8 00:41:37 2024
Wed May  8 00:42:50 2024  n = 500
Wed May  8 00:45:12 2024  n = 1000
Wed May  8 00:47:34 2024  n = 1500
Wed May  8 00:49:53 2024  n = 2000
Wed May  8 00:53:29 2024  n = 2500
Wed May  8 00:55:59 2024  n = 3000
*** OpenAlex2Pajek / Cite - Data Collected Wed May  8 00:56:18 2024
hits: 3086 works: 28206
>>> Citation Cite
>>> publication year
>>> type of publication
>>> language of publication
>>> cited by count
>>> countries distinct count
>>> referenced works
*** OpenAlex2Pajek / Cite - Stop Wed May  8 00:57:04 2024
> # cite8Ci.net: Pajek + TextPad / deg>4 -> VB8.csv
> joinLists("VB7u.csv","VB8.csv","VB8u.csv")
OnN = 443  O\N = 2392  N\O = 14  OUN = 2849
```

```
> Q <- list(
+   search="handball",
+   # filter="publication_year:1980-2023",
+   select=selAll,
+   per_page="200"
+ )
```

Networks

To create the collection we first change the parameter `select` to `selAll`

```
> Q <- list(
+   filter="author.id:A5001676164",
+   select=selAll,
+   per_page="200"
+ )
```

and afterward, run the function `OpenAlex2PajekAll`

```
> OpenAlex2PajekAll(Q,name="batagelj",listF="VB8u.csv")
OpenAlex2Pajek / All - Start Wed May  8 03:50:51 2024
*** OpenAlex2Pajek / All - Process Wed May  8 03:50:51 2024
Wed May  8 03:51:46 2024  n = 500
Wed May  8 03:54:02 2024  n = 1000
Wed May  8 03:56:23 2024  n = 1500
Wed May  8 03:58:49 2024  n = 2000
Wed May  8 04:01:57 2024  n = 2500
Wed May  8 04:04:28 2024  n = 3000
*** OpenAlex2Pajek / All - Data Collected Wed May  8 04:04:58 2024
hits: 3100 works: 28287 authors: 4265 anon: 38 sources: 563
>>> Citation Cite
>>> publication year
>>> type of publication
>>> language of publication
>>> cited by count
>>> countries distinct count
>>> referenced works
>>> Authorship WA
>>> Sources WJ
>>> Keywords WK
>>> Countries WC
*** OpenAlex2Pajek / All - Stop Wed May  8 04:06:00 2024
>
```
