

# The Interdisciplinarity of Economics

## Technical Appendix

Alexandre Truc, Olivier Santerre, Yves Gingras, François Claveau

July 2023

## Contents

<b>1</b>	<b>Preparatory work</b>	<b>1</b>
<b>2</b>	<b>Information on the corpus</b>	<b>8</b>
<b>3</b>	<b>A note on the discourse about interdisciplinarity</b>	<b>14</b>
<b>4</b>	<b>The insularity of economics</b>	<b>16</b>
4.1	Extradisciplinary citations . . . . .	16
4.2	Internal hierarchy (and reproducing Fourcade et al. 2015) . . . . .	33
4.3	Citations toward disciplines . . . . .	42
4.4	The balance of toward and outward citations . . . . .	47
<b>5</b>	<b>Disaggregating the relations of economics</b>	<b>51</b>
5.1	SSH and NSE . . . . .	51
5.2	Strong ties of economics . . . . .	54
5.3	Weak ties of economics . . . . .	77
<b>6</b>	<b>Comparison with the recent literature</b>	<b>83</b>

This document is the accompanying technical appendix to the article of the same name to be published in the *Cambridge Journal of Economics*. The R code used for our analysis is embedded in this file.

## 1 Preparatory work

We load relevant packages:

```
require(data.table)
require(ggplot2)
require(tidyr)
# require(plyr)
require(knitr)
require(RMySQL)
# require(ineq)
# require(grid)
require(dplyr)
# require(data.tree)
require(ggrepel)
require(xtable)
```

Unfortunately, we cannot share our data because they are proprietary (from Clarivate's Web of Science). Most of our analysis will use a data table with 79 million lines, where each line corresponds to one reference having a SSH article either as the citing or the cited document. Furthermore, we have a MySQL database with more data. We define our connection to this MySQL database here:

```
source("connection_info.R") # file with the connection info (usr and pswd)

ESH <- dbConnect(MySQL(), user=usr, password=pswd, dbname='OST_Expanded_SciHum',
                 host='127.0.0.1')
```

We create our disciplinary categories based on the NSF classification while aggregating some fields:

```
# Discipline names:
discipline_info <-
  fread(paste0(
    path_to_discipline_info,
    "Liste_Discipline.txt"))

discipline_info <- discipline_info[,list(Code_Discipline, EGrande_Discipline,
                                       EDiscipline, ESpecialite)]

# We create a new column aggregating some specialties into disciplines
# Everything that is Psychology and below is
# aggregated from the specialty to the Discipline:
i <- discipline_info[EDiscipline=="Psychology", max (Code_Discipline)]
discipline_info[Code_Discipline <= i, discipline := EDiscipline]

# Everything that is Humanities is aggregated
i <- discipline_info[EDiscipline=="Humanities", max (Code_Discipline)]
discipline_info[Code_Discipline <= i, discipline := EDiscipline]

#Everything that is Arts is aggregated
i <- discipline_info[EDiscipline=="Arts", max (Code_Discipline)]
discipline_info[Code_Discipline <= i, discipline := EDiscipline]

#Everything that is Earth and Space is aggregated
i <- discipline_info[EDiscipline=="Earth and Space", max (Code_Discipline)]
discipline_info[Code_Discipline <= i, discipline := EDiscipline]

#Everything that is Biology is aggregated
i <- discipline_info[EDiscipline=="Biology", max (Code_Discipline)]
discipline_info[Code_Discipline <= i, discipline := EDiscipline]

#Everything that is Mathematics is aggregated
i <- discipline_info[EDiscipline=="Mathematics", max (Code_Discipline)]
discipline_info[Code_Discipline <= i, discipline := EDiscipline]

#Everything that is Chemistry is aggregated
i <- discipline_info[EDiscipline=="Chemistry", max (Code_Discipline)]
discipline_info[Code_Discipline <= i, discipline := EDiscipline]

#Everything that is Physics is aggregated
i <- discipline_info[EDiscipline=="Physics", max (Code_Discipline)]
discipline_info[Code_Discipline <= i, discipline := EDiscipline]
```

```

core_socsci <- c("Economics", "Geography", "Political Science and Public Administration",
               "Sociology", "Demography", "International Relations",
               "Anthropology and Archaeology")
core_pro_fie <- c("Management", "Education", "Law")
core_med <- c("General & Internal Medicine", "Psychiatry",
             "Environmental & Occupational Health", "Neurology & Neurosurgery")
cor_eng_tec <- c("Computers", "Operations Research")
#Aggregating social sciences
discipline_info[EDiscipline == "Social Sciences" & !ESpecialite %in% core_socsci,
               discipline := "Other Social Sciences"]
discipline_info[EDiscipline == "Social Sciences" & ESpecialite %in% core_socsci,
               discipline := ESpecialite]

#Aggregating professional fields
discipline_info[EDiscipline == "Professional Fields" & !ESpecialite %in% core_pro_fie,
               discipline := "Other Professional Fields"]
discipline_info[EDiscipline == "Professional Fields" & ESpecialite %in% core_pro_fie,
               discipline := ESpecialite]

#Aggregating clinical medicine
discipline_info[EDiscipline == "Clinical Medicine" & !ESpecialite %in% core_med,
               discipline := "Other Clinical Medicine"]
discipline_info[EDiscipline == "Clinical Medicine" & ESpecialite %in% core_med,
               discipline := ESpecialite]

#Aggregating Engineering and Technology
discipline_info[EDiscipline == "Engineering and Technology" &
               !ESpecialite %in% cor_eng_tec,
               discipline := "Other Engineering \n and Technology"]
discipline_info[EDiscipline == "Engineering and Technology" &
               ESpecialite %in% cor_eng_tec,
               discipline := "Computers & \n Operations Research"]
# all the rest is not aggregate and the ESpecialite becomes the discipline:
discipline_info[is.na(discipline), discipline := EDiscipline]

# Playing with some names to put on two lines:

#Cleaning Political Science name
discipline_info[discipline == "Political Science and Public Administration",
               discipline := "Political Science &\nPublic Administration"]
#Cleaning Anthropology name
discipline_info[discipline == "Anthropology and Archaeology",
               discipline := "Anthropology &\n Archaeology"]

# Core social science and humanities:
core_SSH <- c("Psychology", "Economics",
             "Political Science &\nPublic Administration",
             "Sociology", "Management",
             "Anthropology &\n Archaeology")

# Colors for every disciplines.
# We used Angrist's colors and added purple for Management
colors <- c("Psychology" = "#0000FF",

```

```

"Economics" = "black",
"Sociology" = "#FF0000",
"Anthropology &\n Archaeology" = "#CCCC00", "Management" = "purple",
"Political Science &\nPublic Administration" = "#228B22",
"Others" = alpha("gray",.3)

```

*# Linetypes for every disciplines. We used Angrist's limetype and  
# added dotdash for Management*

```

linetypes <- c("Psychology" = "longdash",
              "Economics" = "solid",
              "Sociology" = "dotted",
              "Anthropology &\n Archaeology" = "solid",
              "Political Science &\nPublic Administration" = "dashed",
              "Management" = "dotdash",
              "Others" = "solid")

```

We regroup most functions used later in the following code block:

*#Renaming Anthropology and Political Science to match Angrist et al's discipline names*

```

renaming_ssh_angrist <- function(dt_citing_cited_journals){
  dt_citing_cited_journals[Discipline_citing == "Anthropology &\n Archaeology",
                           Discipline_citing := "Anthropology"]
  dt_citing_cited_journals[Discipline_cited == "Anthropology &\n Archaeology",
                           Discipline_cited := "Anthropology"]
  dt_citing_cited_journals[
    Discipline_citing == "Political Science &\nPublic Administration",
    Discipline_citing := "Political Science"]
  dt_citing_cited_journals[
    Discipline_cited == "Political Science &\nPublic Administration",
    Discipline_cited := "Political Science"]
  dt_citing_cited_journals
}

```

*#Recategorizing journals to match Angrist et al's categories*

```

renaming_business_angrist <- function(dt_citing_cited_journals, dt_angrist_journals){
  dt_citing_cited_journals[citing_code_revue %in%
                           dt_angrist_journals[angrist_disc == "Accounting"]$Code_Revue,
                           `:=` (citing_discipline = 201, Discipline_citing =
  dt_citing_cited_journals[cited_code_revue %in%
                           dt_angrist_journals[angrist_disc == "Accounting"]$Code_Revue,
                           `:=` (cited_discipline = 201, Discipline_cited =
                               "Accounting")]
  dt_citing_cited_journals[citing_code_revue %in%
                           dt_angrist_journals[angrist_disc == "Finance"]$Code_Revue,
                           `:=` (citing_discipline = 202,
                               Discipline_citing = "Finance")]
  dt_citing_cited_journals[cited_code_revue %in%
                           dt_angrist_journals[angrist_disc == "Finance"]$Code_Revue,
                           `:=` (cited_discipline = 202,
                               Discipline_cited = "Finance")]
  dt_citing_cited_journals[citing_code_revue %in%
                           dt_angrist_journals[angrist_disc == "Marketing"]$Code_Revue,

```

```

      `:=` (citing_discipline = 203,
           Discipline_citing = "Marketing")]
dt_citing_cited_journals[cited_code_revue %in%
      dt_angrist_journals[angrist_disc == "Marketing"]$Code_Revue,
      `:=` (cited_discipline = 203,
           Discipline_cited = "Marketing")]
dt_citing_cited_journals
}

adding_disc_name_to_dt_citing_cited <- function(dt_citing_cited_all)
{
  dt_citing_cited_all <- merge(dt_citing_cited_all,
                             discipline_info[,list(Code_Discipline, discipline)],
                             by.x = "citing_discipline", by.y = "Code_Discipline")
  setnames(dt_citing_cited_all, "discipline", "Discipline_citing")
  dt_citing_cited_all <- merge(dt_citing_cited_all,
                             discipline_info[,list(Code_Discipline, discipline)],
                             by.x = "cited_discipline", by.y = "Code_Discipline")
  setnames(dt_citing_cited_all, "discipline", "Discipline_cited")
  dt_citing_cited_all
}

# Set of functions adapted from
# https://gist.github.com/eliocamp/eabafab2825779b88905954d84c82b32
# on 2019-11-14
# They are used to generate two linetype aesthetics
# in the same graph with separate legends:
#' Allows to add another scale
#'
#' @param new_aes character with the aesthetic for which new scales will be
#' created
#'
new_scale <- function(new_aes) {
  structure(ggplot2::standardise_aes_names(new_aes), class = "new_aes")
}

#' Convenient functions
new_scale_fill <- function() {
  new_scale("fill")
}

new_scale_color <- function() {
  new_scale("colour")
}

new_scale_colour <- function() {
  new_scale("colour")
}

new_scale_linetype <- function() {
  new_scale("linetype")
}

```

```

#' Special behaviour of the "+" for adding a `new_aes` object
#' It changes the name of the aesthetic for the previous layers, appending
#' "_new" to them.
ggplot_add.new_aes <- function(object, plot, object_name) {
  plot$layers <- lapply(plot$layers, bump_aes, new_aes = object)
  plot$scales$scales <- lapply(plot$scales$scales, bump_aes, new_aes = object)
  plot$labels <- bump_aes(plot$labels, new_aes = object)
  plot
}

bump_aes <- function(layer, new_aes) {
  UseMethod("bump_aes")
}

bump_aes.Scale <- function(layer, new_aes) {
  old_aes <- layer$aesthetics[remove_new(layer$aesthetics) %in% new_aes]
  new_aes <- paste0(old_aes, "_new")

  layer$aesthetics[layer$aesthetics %in% old_aes] <- new_aes

  if (is.character(layer$guide)) {
    layer$guide <- match.fun(paste("guide_", layer$guide, sep = ""))()
  }
  layer$guide$available_aes[layer$guide$available_aes %in% old_aes] <- new_aes
  layer
}

bump_aes.Layer <- function(layer, new_aes) {
  original_aes <- new_aes

  old_aes <- names(layer$mapping)[remove_new(names(layer$mapping)) %in% new_aes]
  new_aes <- paste0(old_aes, "_new")

  old_geom <- layer$geom

  old_setup <- old_geom$handle_na
  new_setup <- function(self, data, params) {
    colnames(data)[colnames(data) %in% new_aes] <- original_aes
    old_setup(data, params)
  }

  new_geom <- ggplot2::ggproto(paste0("New", class(old_geom)[1]), old_geom,
                              handle_na = new_setup)

  new_geom$default_aes <- change_name(new_geom$default_aes, old_aes, new_aes)
  new_geom$non_missing_aes <- change_name(new_geom$non_missing_aes, old_aes, new_aes)
  new_geom$required_aes <- change_name(new_geom$required_aes, old_aes, new_aes)
  new_geom$optional_aes <- change_name(new_geom$optional_aes, old_aes, new_aes)

  layer$geom <- new_geom

  old_stat <- layer$stat

```

```

old_setup2 <- old_stat$handle_na
new_setup <- function(self, data, params) {
  colnames(data)[colnames(data) %in% new_aes] <- original_aes
  old_setup2(data, params)
}

new_stat <- ggplot2::ggproto(paste0("New", class(old_stat)[1]), old_stat,
  handle_na = new_setup)

new_stat$default_aes <- change_name(new_stat$default_aes, old_aes, new_aes)
new_stat$non_missing_aes <- change_name(new_stat$non_missing_aes, old_aes, new_aes)
new_stat$required_aes <- change_name(new_stat$required_aes, old_aes, new_aes)
new_stat$optional_aes <- change_name(new_stat$optional_aes, old_aes, new_aes)

layer$stat <- new_stat

layer$mapping <- change_name(layer$mapping, old_aes, new_aes)
layer
}

bump_aes.list <- function(layer, new_aes) {
  old_aes <- names(layer)[remove_new(names(layer)) %in% new_aes]
  new_aes <- paste0(old_aes, "_new")

  names(layer)[names(layer) %in% old_aes] <- new_aes
  layer
}

change_name <- function(list, old, new) {
  UseMethod("change_name")
}

change_name.character <- function(list, old, new) {
  list[list %in% old] <- new
  list
}

change_name.default <- function(list, old, new) {
  nam <- names(list)
  nam[nam %in% old] <- new
  names(list) <- nam
  list
}

change_name.NULL <- function(list, old, new) {
  NULL
}

remove_new <- function(aes) {
  stringi::stri_replace_all(aes, "", regex = "(new)*")
}

```

Finally, a few parameters to be reused later:

```

min_period = 1950; max_period = 2018

# Number disciplines in graph:
n_disc = 7

# To change the period on which we count the most cited disciplines
firstYearImpDisc = 2011

```

## 2 Information on the corpus

We first want general information on the journals, starting with the number of journals per disciplines in the SSH:

```

# loading data
dt_citing_cited <- readRDS(paste0(path_to_project_data,
                                  "dt_citing_cited_all.rds"))

# aggregating by SSH disciplines
dt_journals_disc <- dt_citing_cited[,
                                     list(ID_Art, Year, citing_discipline,
                                           citing_code_revue)] %>%
  unique() %>% .[Year %in% min_period:(max_period) & citing_discipline>100] %>%
  .[,list(tot_j=length(unique(citing_code_revue))) ,
      by=citing_discipline]

dt_journals_disc <- merge(dt_journals_disc,
                         discipline_info[,list(Code_Discipline,discipline)],
                         by.x="citing_discipline", by.y="Code_Discipline")

dt_journals_disc <- dt_journals_disc[citing_discipline<200,
                                     list(Number_of_journals =sum(tot_j)),
                                     by=discipline]

# Note taht "<200" is there to remove "Unknown"

dt_journals_disc <- dt_journals_disc[order(-Number_of_journals)]

pos_core_ssh <- which(dt_journals_disc$discipline %in% core_SSH)
pos_other_ssh <- which(! 1:nrow(dt_journals_disc) %in% pos_core_ssh)

dt_journals_disc[c(pos_core_ssh,pos_other_ssh)]

```

```

##              discipline Number_of_journals
## 1:              Psychology              721
## 2:              Management              459
## 3:              Economics              384
## 4:      Anthropology &\n Archaeology        219
## 5: Political Science &\nPublic Administration    195
## 6:              Sociology              160
## 7:              Humanities             1760
## 8:              Health                632

```



```

## 9:                Other Social Sciences                608
## 10:                Arts                                390
## 11:                Education                            364
## 12:                Other Professional Fields            342
## 13:                Law                                 232
## 14:                Geography                            112
## 15:                International Relations              88
## 16:                Demography                          33

# Saving journal numbers
tot_j_ssh <- dt_journals_disc[discipline != "Unknown",Number_of_journals] %>% sum()
# also for NSE journals:
n_journals_NSE <- c(
  dt_citing_cited[Year %in% min_period:(max_period) &
    citing_discipline<=100,unique(citing_code_revue)] ,

  dt_citing_cited[Year %in% min_period:(max_period) &
    cited_discipline<=100,unique(cited_code_revue)] ) %>%
  unique() %>% length()

rm(dt_citing_cited,dt_journals_disc)

```

The total number of journals in the SSH is 6699. Their number increases through time as the following figure shows:

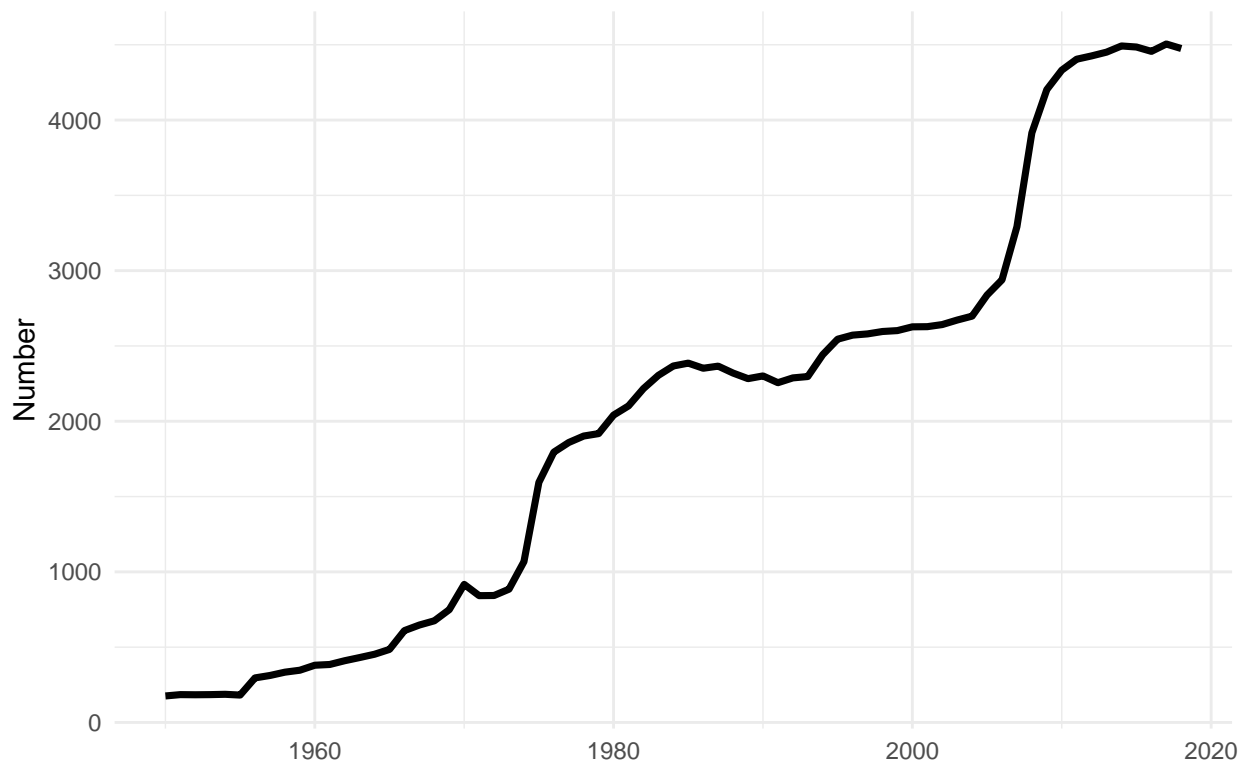
```

dt_citing_cited <- readRDS(paste0(path_to_project_data,
  "dt_citing_cited_all.rds"))

#Plotting the number of journals per year in the corpus
dt_citing_cited <- dt_citing_cited[between(Year, min_period, max_period),]
dt_citing_cited <- dt_citing_cited[between(citing_discipline,101,200),
  list(ID_Art,Year, citing_discipline,
    citing_code_revue)]
nb_journals <- dt_citing_cited[, list(Year, citing_code_revue)] %>% unique()
nb_journals <- nb_journals[, rev_total := .N, by = Year]
ggplot(nb_journals, aes(x=Year, y=rev_total)) + geom_line(lwd=1.25) + theme_minimal() +
  xlab("") + ylab("Number") +
  ggtitle(
    "Number of SSH journals per year in the corpus"
  )

```

## Number of SSH journals per year in the corpus



```
rm(dt_citing_cited)
```

There are also 13517 journals from the NSE that cite at least once the SSH or are cited at least once by the SSH.

Next, we look at the distribution of articles in the SSH:

```
dt_citing_cited <- readRDS(paste0(path_to_project_data,
                                  "dt_citing_cited_all.rds"))
dt_citing_cited <- dt_citing_cited[between(Year, min_period, max_period),]

# Two main data table, one with every articles and their references (dt_citing_cited)
# and one with every articles only (dt_all_art)
dt_citing_cited <- dt_citing_cited[between(citing_discipline,101,200),
                                   list(ID_Art,Year,
                                         citing_discipline,citing_code_revue)]
dt_citing_cited <- merge(dt_citing_cited,
                        discipline_info[,list(Code_Discipline,discipline)],
                        by.x = "citing_discipline", by.y= "Code_Discipline")
dt_all_art <- unique(dt_citing_cited)

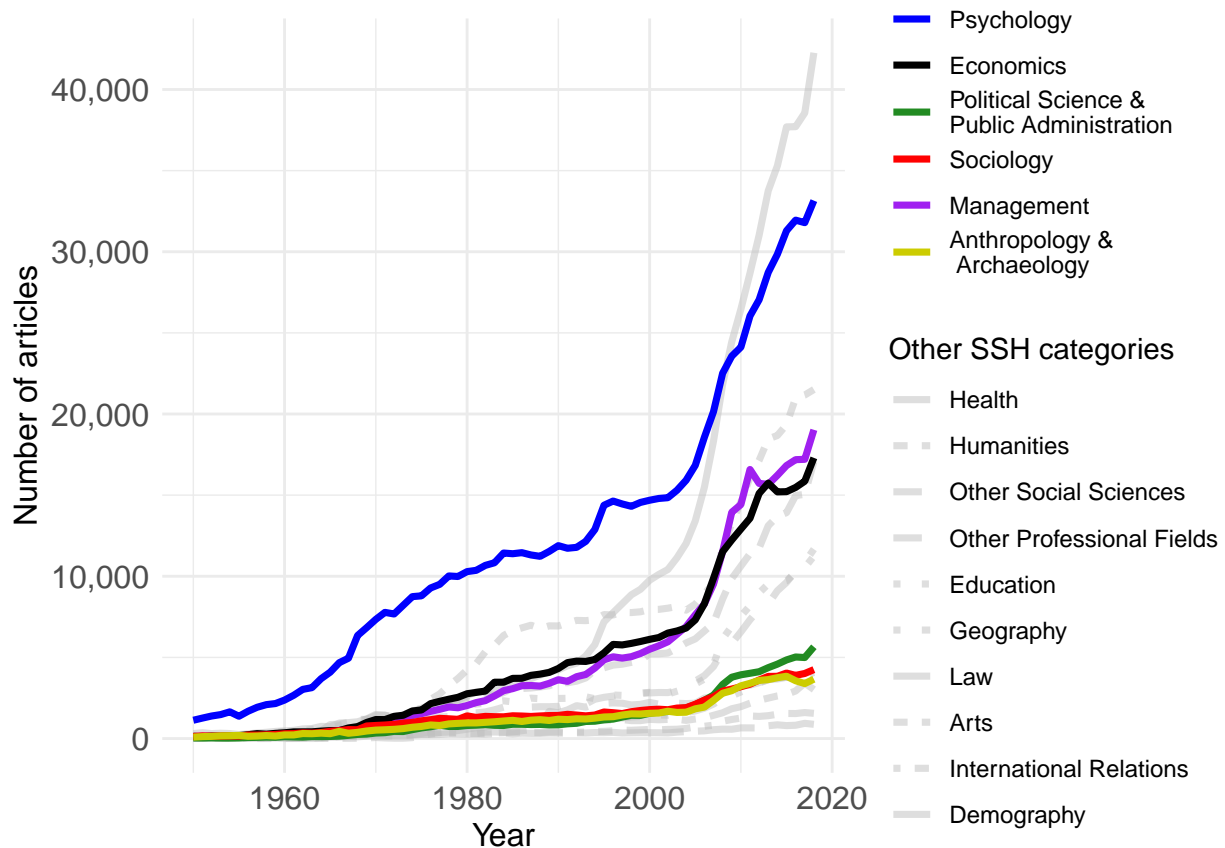
#Counting number of article per discipline per year
dt_art_by_j_y <- dt_all_art[,.(nb_art=.N), by= .(Year,discipline)]
setkey(dt_art_by_j_y,discipline,Year)

# Finding the order at the end to order in the legend
order_disc <- dt_art_by_j_y[Year== max_period][order(-nb_art),discipline]
dt_art_by_j_y$discipline <- factor(dt_art_by_j_y$discipline, levels = order_disc)
```

```

#Ploting number of article per discipline and per year in corpus
ggplot() + geom_line(dt_art_by_j_y[!discipline %in% core_SSH],
                    mapping = aes(x=Year, y = nb_art, linetype = discipline),
                    colour = alpha("gray",.5),lwd=1.25) +
geom_line(dt_art_by_j_y[discipline %in% core_SSH],
          mapping = aes(x=Year, y=nb_art, col = discipline), lwd=1.25) +
scale_color_manual(name = "Main social sciences", breaks = core_SSH, values = colors) +
scale_linetype_discrete(name="Other SSH categories") +
scale_y_continuous(labels=function(x) format(x, big.mark = ",", scientific = FALSE)) +
guides(color = guide_legend(order = 1),
        linetype = guide_legend(order = 2)) + ylab("Number of articles") +
theme_minimal() + theme(axis.text = element_text(size=12), axis.title = element_text(size=12))

```



```

ggsave( paste0("figures/", "fig1.png"), width = 16, height = 9, bg = "white",
        scale = .6)

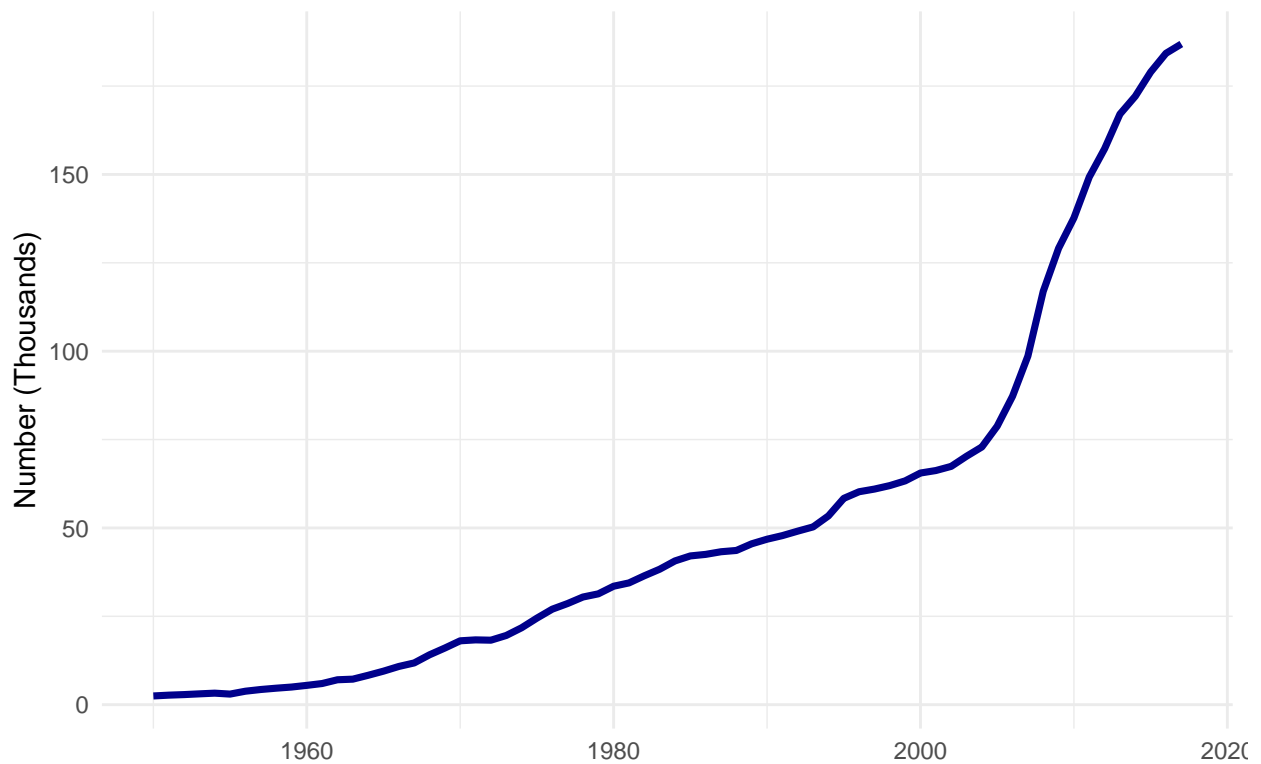
```

```

#Ploting number of article total per year in corpus
nb_art_year <- dt_art_by_j_y[, tot_article := sum(nb_art), by=Year]
ggplot(nb_art_year[Year<max_period], aes(x=Year, tot_article/1000)) +
geom_line(lwd=1.25,col="darkblue") + theme_minimal() +
ylab("Number (Thousands)") + xlab("") +
ggtitle("Number of SSH articles per year in the corpus")

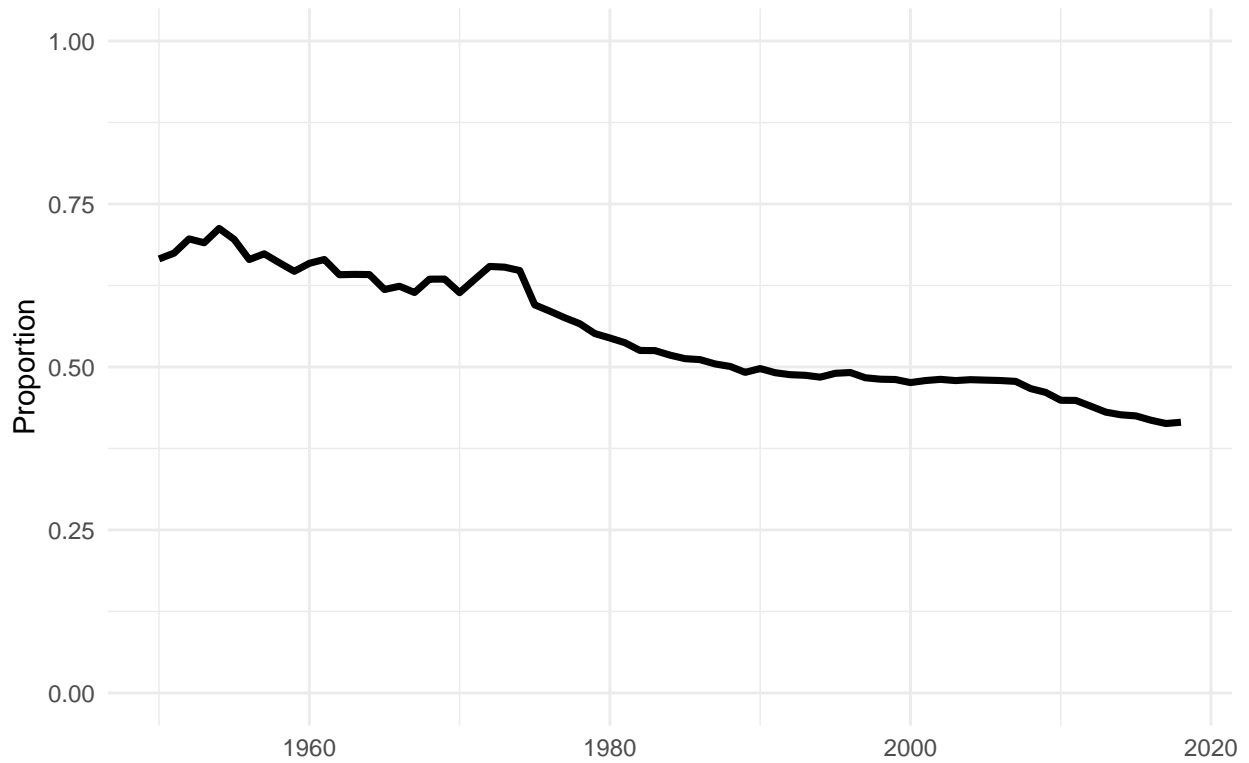
```

Number of SSH articles per year in the corpus



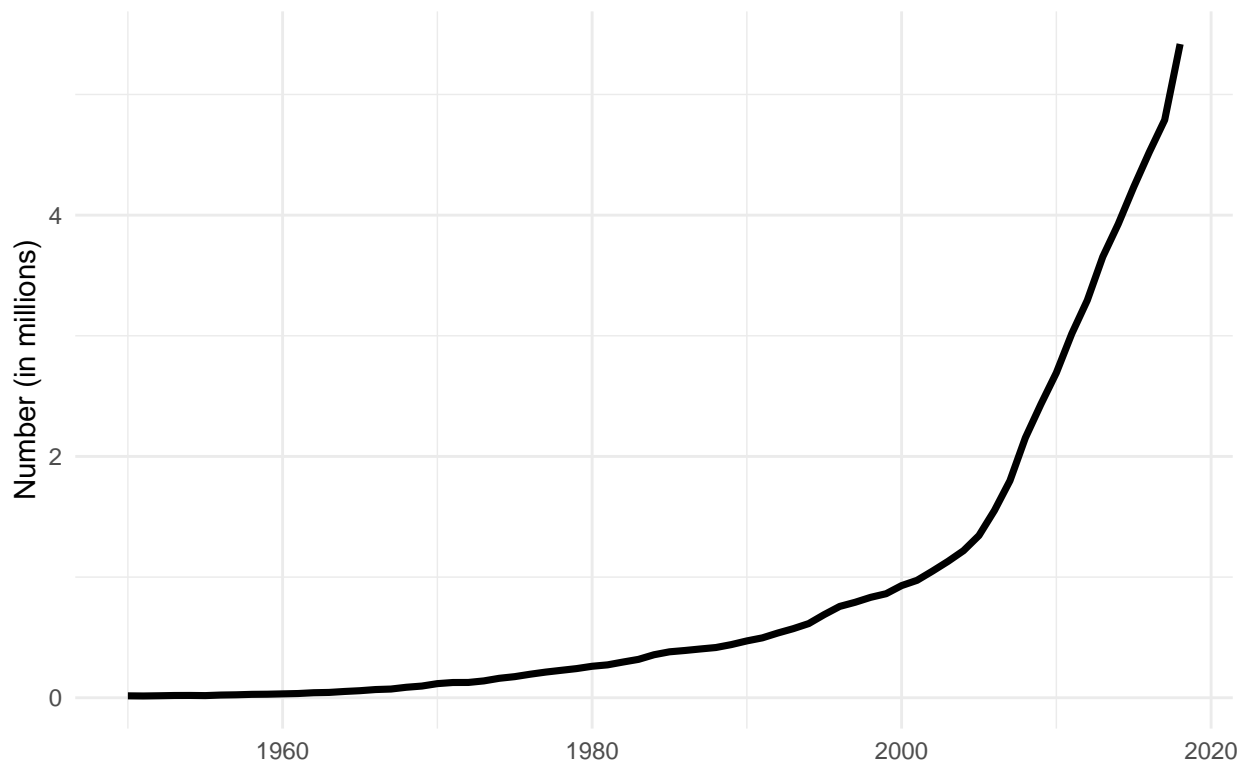
```
#Plotting proportion of articles in the main SSH discipline per year
prop_ssh <- nb_art_year[discipline %in% core_SSH,
  tot_ssh := sum(nb_art), by = Year
][,
  list(Proportion = tot_ssh/tot_article,
    Year)][!is.na(Proportion),] %>%
  unique()
ggplot(prop_ssh, aes(x=Year, y= Proportion)) + geom_line(lwd=1.25) + theme_minimal() +
  ylim(0,1) + xlab("") + ylab("Proportion") +
  ggtitle("Proportion of articles in the main SSH disciplines per year")
```

Proportion of articles in the main SSH disciplines per year



```
#Plotting number of references per year in corpus  
nb_ref <- dt_citing_cited[, tot_ref := .N, by = Year][,list(Year, tot_ref)] %>% unique()  
ggplot(nb_ref, aes(x=Year, y = tot_ref/1000000)) + geom_line(lwd=1.25) +  
  theme_minimal() + xlab("") + ylab("Number (in millions)") +  
  ggtitle("References per year in the corpus")
```

## References per year in the corpus



```
rm(dt_citing_cited,dt_art_by_j_y,nb_ref,nb_art_year,dt_all_art)
```

### 3 A note on the discourse about interdisciplinarity

Larivière and Gingras (2014, 188) distinguish between the discourse and the practice of interdisciplinarity. Their figure for the discourse looks at the proportion of articles with “interdisciplinary” in the title. We check whether we can reproduce the graph with our data, focusing on SSH articles:

```
# The articles with one word referring to inter/trans/multi -disciplinarity in title
interdisc_discourse <-
dbGetQuery(ESH,
paste0("SELECT Annee_Bibliographique as Year, ID_Art,Code_Revue
FROM OST_Expanded_SciHum.Articles
WHERE Titre LIKE '%INTER-DISCIPLIN%' OR
Titre LIKE '%INTERDISCIPLIN%' OR
Titre LIKE '%TRANS-DISCIPLIN%' OR
Titre LIKE '%TRANSDISCIPLIN%' OR
Titre LIKE '%MULTI-DISCIPLIN%' OR
Titre LIKE '%MULTIDISCIPLIN%'
ORDER BY Annee_Bibliographique ;")) %>%
data.table
# Our period of reference:
interdisc_discourse <- interdisc_discourse[Year %in% min_period:2015]
# We use the number of articles per year computed in earlier code chunk
# Counting per years:
inter_per_y <- interdisc_discourse[,list(nb_inter = .N),by=Year]
```

```

# Getting total number of articles per year
nb_art_year <-
dbGetQuery(ESH, paste0(
"SELECT Annee_Bibliographique as Year, count(ID_Art) as tot_article
FROM OST_Expanded_SciHum.Articles
group by Annee_Bibliographique
order by Annee_Bibliographique;")) %>%
  data.table
# Our period of reference:
nb_art_year <- nb_art_year[Year %in% min_period:2015]

# Merging tables on year and computing ratio
setkey(inter_per_y, Year); setkey(nb_art_year,Year)

inter_per_y <- inter_per_y[nb_art_year] ; rm(nb_art_year)

inter_per_y[,prop_interdisc:= nb_inter/tot_article]

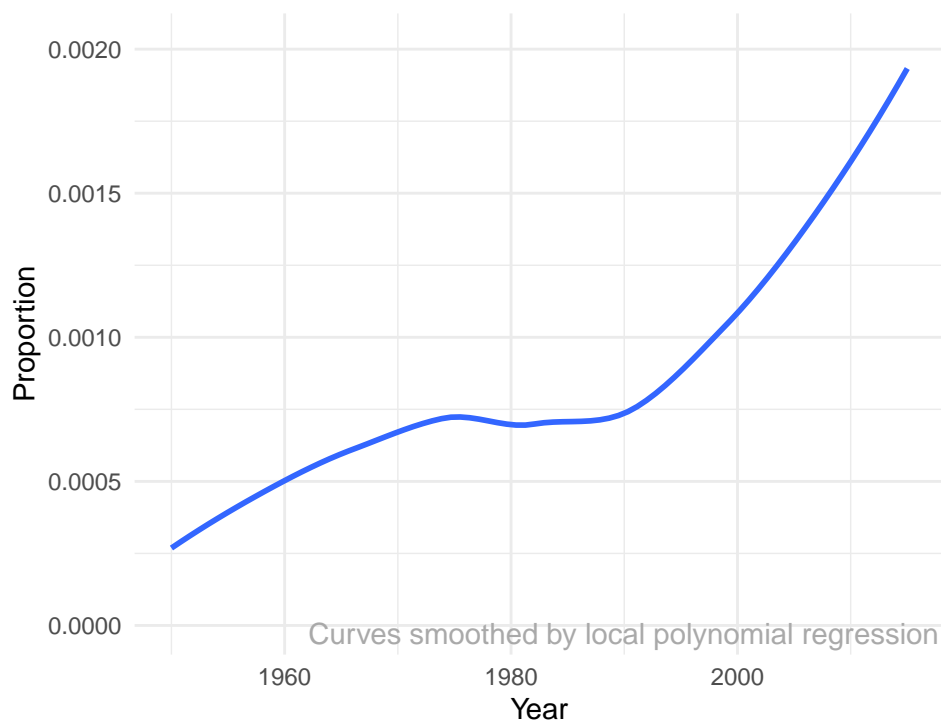
#ggplot(data = interdisc_discourse, mapping = aes(x=Year,y=prop_interdisc)) + geom_line()

#plot(interdisc_discourse[,list(Year,prop_interdisc)])
ggplot(inter_per_y, aes(x=Year, y=prop_interdisc)) +
  geom_smooth(se=F, method = 'loess') +
  theme_minimal() + ylab("Proportion") + ylim(0,max(inter_per_y$prop_interdisc)) +
  ggtitle(
    "Proportion of articles in the SSH with titles explicitly\n
    referring to inter/trans/multi-disciplinarity") +
  annotate( "text",x=max(inter_per_y$Year),y=-Inf,hjust=.95,vjust=-0.5,
    label="Curves smoothed by local polynomial regression",
    size=4,color="darkgrey" )

## `geom_smooth()` using formula = 'y ~ x'

```

## Proportion of articles in the SSH with titles explicitly referring to inter/trans/multi-disciplinarity



We get the same pattern as Larivière and Gingras on the plateau of popularity of the subject in the 1980s and the extent to which it is growing ever more common recently.

## 4 The insularity of economics

### 4.1 Extradisciplinary citations

We turn to the results used in the article: a measure of aggregate interdisciplinarity *in practice*, not in the discourse. Our main results in the article use citations without weighing the journal where the citations originate. The code in the next subsection covers this case of unweighed citations. The second subsection reproduces exactly the same analysis, but weighs citations by a measure of journal impact. The third subsection splits disciplines between top journals and the rest.

#### 4.1.1 Unweighed citations

We build our main object with citing and cited discipline in all years.

```
dt_citing_cited <- readRDS(paste0(path_to_project_data,
                                "dt_citing_cited_all.rds"))
dt_citing_cited <- dt_citing_cited[between(Year, 1950, max_period),]
#Summing how many citation goes from every dicipline to every other dicipline per year
citing_cited_disc <- dt_citing_cited[, list(Citations =.N),
                                     by= c("Year", "citing_discipline",
                                             "cited_discipline")]
#Merging with dicipline_info to get cited dicipline names
```



```

citing_cited_disc <- merge(citing_cited_disc,
                           discipline_info[,list(EGrande_Discipline, Code_Discipline,
                                                  Discipline_cited = discipline)],
                           by.x="cited_discipline", by.y="Code_Discipline", all.x=T)
#Merging with dicipline_info to get citing discipline names
citing_cited_disc <- merge(citing_cited_disc,
                           discipline_info[,list(Code_Discipline,
                                                  Discipline_citing = discipline)],
                           by.x="citing_discipline", by.y="Code_Discipline", all.x=T)
rm(dt_citing_cited)

```

First, an aggregate measure of the degree of interdisciplinarity of our main SSH disciplines (COC or Citations Outside Category):

```

# Aggregating by big disciplines instead of the disaggregated ones:

```

```

citing_cited_bigdisc <- citing_cited_disc[,
                                           list(Citations= sum(Citations)),

```

```

# Focussing on core SSH and aggregating at the year level

```

```

interdisc_per_y <- citing_cited_bigdisc[
  !is.na(Discipline_cited) &
  Discipline_citing %in% core_SSH,
  list(Proportion = 1-
       .SD[Discipline_citing==Discipline_cited,sum(Citations)]
  by= Year]

```

```

setkey(interdisc_per_y,Year)

```

```

ggplot(interdisc_per_y,aes(x=Year,y=Proportion)) + #geom_line() + #

```

```

  geom_smooth(se=F, method="loess")+
  ylim(0,max(interdisc_per_y$Proportion)) + theme_minimal() +
  theme(axis.text=element_text(size=10),
        axis.title=element_text(size=12#,face="bold
        )) +

```

```

  annotate( "text", x=max(interdisc_per_y$Year),
           y=-Inf,hjust=1,vjust=-0.5,
           label="Curve smoothed by local polynomial regression",
           size=4,color="darkgrey" ) +

```

```

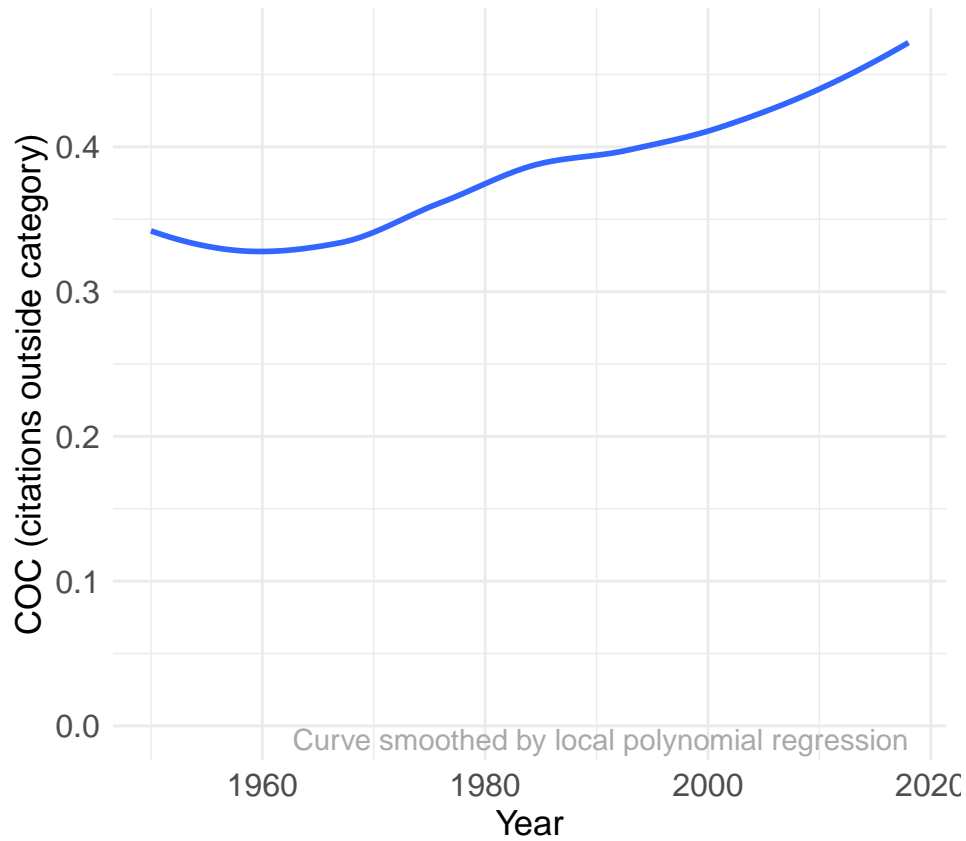
  ylab("COC (citations outside category)") + theme(axis.text = element_text(size=12), axis.title = el

```

```

## `geom_smooth()` using formula = 'y ~ x'

```



```
ggsave( paste0("figures/", "fig3.png"), width = 16, height = 9, bg = "white",
        scale = .6)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

The same data in a table:

```
interdisc_per_y
```

```
##      Year Proportion
## 1: 1950 0.3953729
## 2: 1951 0.3183225
## 3: 1952 0.3361748
## 4: 1953 0.3414831
## 5: 1954 0.3150213
## 6: 1955 0.3401836
## 7: 1956 0.3168027
## 8: 1957 0.3185131
## 9: 1958 0.3187697
## 10: 1959 0.3388829
## 11: 1960 0.3325390
## 12: 1961 0.3056352
## 13: 1962 0.3240866
## 14: 1963 0.3162289
## 15: 1964 0.3177575
## 16: 1965 0.3175595
## 17: 1966 0.3304379
## 18: 1967 0.3273210
## 19: 1968 0.3372873
```

##	20:	1969	0.3287567
##	21:	1970	0.3379434
##	22:	1971	0.3304403
##	23:	1972	0.3499440
##	24:	1973	0.3370341
##	25:	1974	0.3482150
##	26:	1975	0.3598568
##	27:	1976	0.3668135
##	28:	1977	0.3704759
##	29:	1978	0.3713307
##	30:	1979	0.3727372
##	31:	1980	0.3851823
##	32:	1981	0.3832727
##	33:	1982	0.3902215
##	34:	1983	0.3860762
##	35:	1984	0.3936496
##	36:	1985	0.3964173
##	37:	1986	0.3940561
##	38:	1987	0.3988312
##	39:	1988	0.3910752
##	40:	1989	0.3898815
##	41:	1990	0.3936273
##	42:	1991	0.3945068
##	43:	1992	0.3966732
##	44:	1993	0.3845597
##	45:	1994	0.3917470
##	46:	1995	0.4020827
##	47:	1996	0.4008529
##	48:	1997	0.4049961
##	49:	1998	0.4112799
##	50:	1999	0.4061033
##	51:	2000	0.4059220
##	52:	2001	0.4059653
##	53:	2002	0.4070790
##	54:	2003	0.4124841
##	55:	2004	0.4140607
##	56:	2005	0.4234574
##	57:	2006	0.4260757
##	58:	2007	0.4347069
##	59:	2008	0.4371863
##	60:	2009	0.4403649
##	61:	2010	0.4380028
##	62:	2011	0.4427045
##	63:	2012	0.4484406
##	64:	2013	0.4536170
##	65:	2014	0.4532994
##	66:	2015	0.4630352
##	67:	2016	0.4628947
##	68:	2017	0.4661130
##	69:	2018	0.4721364
##		Year	Proportion

Second, we take a closer look at the COC for every core SSH disciplines.

The following graph shows the total COC through time for each of our core SSH disciplines, other SSH

disciplines are in gray. The next graph is a monochrome version of the same figure where we focus on the core SSH disciplines for readability.

```

dt_citing_cited <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))
dt_citing_cited <- dt_citing_cited[between(Year, 1950, max_period)]
dt_citing_cited <- merge(dt_citing_cited,
                        discipline_info[,list(Code_Discipline,
                                              Discipline_cited = discipline)],
                        by.x="cited_discipline", by.y="Code_Discipline", all.x=T)
dt_citing_cited <- merge(dt_citing_cited,
                        discipline_info[,list(Code_Discipline,
                                              Discipline_citing = discipline)],
                        by.x="citing_discipline", by.y="Code_Discipline", all.x=T)

interdisc_per_y_and_disc <- dt_citing_cited[!is.na(cited_discipline) &
                                           !is.na(citing_discipline)]
interdisc_per_y_and_disc <- unique(
  interdisc_per_y_and_disc[,
    list(Discipline_cited,
         Citations = .N),
        by= list(Year,Discipline_citing, Discipline_cited)])
interdisc_per_y_and_disc <- interdisc_per_y_and_disc[,
  list(Discipline_cited,
       Proportion =
         Citations/sum(Citations)),
        by= list(Year,Discipline_citing)]

setkey(interdisc_per_y_and_disc, Year, Discipline_citing)
# The plot data, focussing on core SSH
dt_plot <- interdisc_per_y_and_disc[Discipline_citing==Discipline_cited,
  list(Year, Discipline_citing,
       Prop = 1-Proportion)]

# Smoothing y coordinate outside ggplot with loess
dt_plot <- dt_plot[,list(Year, Prop,
                       Proportion = predict(loess(Prop~Year), Year)),
  by=Discipline_citing]

# Finding the order at the end to order in the legend
order_disc <- dt_plot[Year== 2016][order(-Proportion),Discipline_citing]
dt_plot$Discipline_citing <- factor(dt_plot$Discipline_citing, levels = order_disc)
dt_plot <- dt_plot[!is.na(Discipline_citing)]
ymin= min(dt_plot$Proportion); ymax =max(dt_plot$Proportion)

#Abbreviations for the labels at the end of the line
dt_plot[Discipline_citing == "Psychology", Disc_abreviation := "Psychology"]
dt_plot[Discipline_citing == "Economics", Disc_abreviation := "Economics"]
dt_plot[Discipline_citing == "Political Science &\nPublic Administration",
  Disc_abreviation := "Political Science"]
dt_plot[Discipline_citing == "Sociology", Disc_abreviation := "Sociology"]
dt_plot[Discipline_citing == "Management", Disc_abreviation := "Management"]
dt_plot[Discipline_citing == "Anthropology &\n Archaeology",
  Disc_abreviation := "Anthropology"]

```

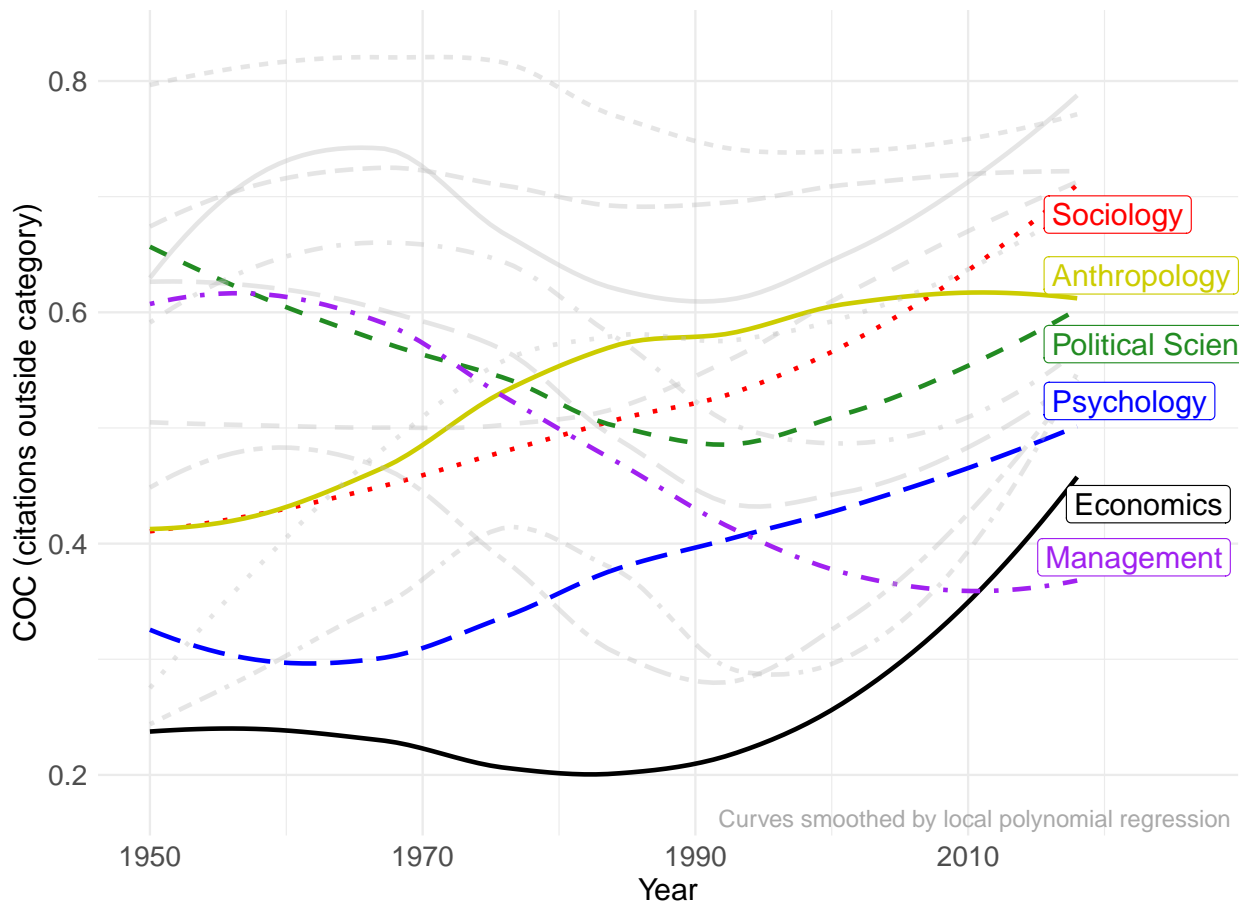
```

dt_plot$Discipline_for_plot <- dt_plot$Discipline_citing
dt_plot[!Discipline_citing %in% core_SSH, Discipline_citing := "Others"]

#Plotting
dt_plot[Discipline_citing %in% core_SSH] %>%
  mutate(label = if_else(Year == max(Year),
                        as.character(Disc_abbreviation), NA_character_)) %>%
  ggplot(mapping = aes(x=Year, y=Prop, col = Discipline_citing)) +
  geom_smooth(se=F,method = 'loess',
            aes(x=Year, y=Prop,
                col = Discipline_citing, linetype = Discipline_citing),
            show.legend=FALSE) +
  scale_color_manual(name = "Disciplines", breaks = core_SSH, values = colors) +
  scale_linetype_manual(name = "Disciplines", breaks = core_SSH, values = linetypes) +
  new_scale_linetype() +
  geom_smooth(dt_plot[Discipline_citing == "Others" & Discipline_for_plot != "Arts"],
            mapping = aes(x=Year, y = Prop#, group = Discipline_for_plot
                          , linetype = Discipline_for_plot),
            color = alpha("gray",.4),
            se=F, method = 'loess') +
  scale_linetype_discrete(name="Other Disciplines") + theme_minimal() +
  theme(legend.position="bottom", legend.box = "vertical") +
  ylab("COC (citations outside category)") +
  xlim(c(NA, 2026)) +
  theme(axis.text=element_text(size=13),
        axis.title=element_text(size=14),
        legend.text=element_text(size=13),
        legend.title=element_text(size=14)
        ) +
  guides(linetype=guide_legend(keywidth = 3, keyheight = 1,nrow=3,byrow=TRUE)) +
  geom_label_repel(aes(label = label), nudge_x = 1, xlim = c(2015, NA),
                  na.rm = TRUE, show.legend = FALSE, size =5) +
  annotate( "text", x=max(dt_plot$Year), y=-Inf,
          hjust=.7, vjust=-0.5,
          label="Curves smoothed by local polynomial regression",
          size=4, color="darkgrey" )

## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'

```



Disciplines

—	Demography	- - - -	Other Social Sciences	- - - -	Health
- - - -	Other Professional Fields	· · · · ·	Geography	· · · · ·	International
- - - -	Education	- - - -	Humanities	· · · · ·	Law

```
ggsave( paste0("figures/", "fig2.png"), width = 16, height = 9, bg = "white",
scale = .7)
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

```
shades_of_gray <- c("Psychology" = "#404040",
                    "Economics" = "black",
                    "Sociology" = "#202020",
                    "Anthropology & \n Archaeology" = "#A0A0A0",
                    "Political Science & \n Public Administration" = "#606060",
                    "Management" = "808080",
                    "Others" = alpha("gray",.3))
```

```
dt_plot[Discipline_citing %in% core_SSH] %>%
  mutate(label = if_else(Year == max(Year),
                        as.character(Disc_abbreviation), NA_character_)) %>%
```

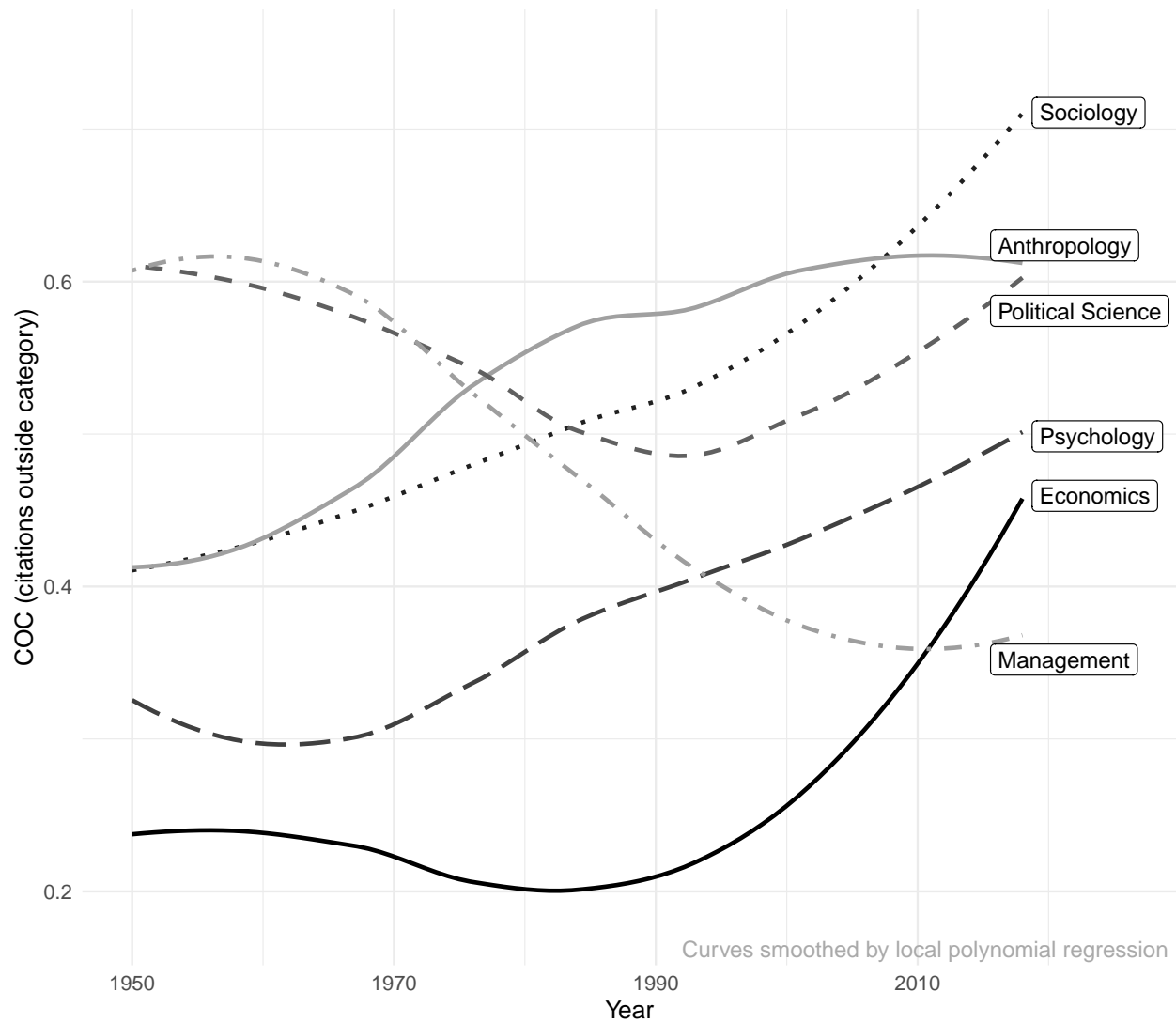
```

ggplot(mapping = aes(x=Year, y=Prop, gr = Discipline_citing)) +
  geom_smooth(se=F,method = 'loess',
             aes(x=Year, y=Prop,col = Discipline_citing,
                 linetype = Discipline_citing),show.legend=FALSE) +
  scale_color_manual(name = "Disciplines", breaks = core_SSH, values = shades_of_gray) +
  scale_linetype_manual(name = "Disciplines", breaks = core_SSH, values = linetypes) +
  theme_minimal() +
  theme(legend.position="bottom", legend.box = "vertical") +
  ylab("COC (citations outside category)") +
  xlim(c(NA, 2026)) + ylim(c(NA, 0.75)) +
  theme(axis.text=element_text(size=10),
        axis.title=element_text(size=12)) +
  geom_label_repel(aes(label = label), nudge_x = 1, xlim = c(2015, NA), na.rm = TRUE,
                  show.legend = FALSE) +
  annotate( "text",x=max(dt_plot$Year),y=-Inf,hjust=.7,vjust=-0.5,
          label="Curves smoothed by local polynomial regression",
          size=4,color="darkgrey" )

```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).
```



```
rm(dt_citing_cited); rm(interdisc_per_y_and_disc)
```

The same data in a table:

```
COC_econ_whole_period <- dt_plot[Discipline_citing == "Economics"]
# Printing the table
dt_plot[,list(Discipline_citing, Year, Prop )]
```

```
##           Discipline_citing Year      Prop
##  1: Anthropology &\n Archaeology 1950 0.4736842
##  2: Anthropology &\n Archaeology 1951 0.4385633
##  3: Anthropology &\n Archaeology 1952 0.4357616
##  4: Anthropology &\n Archaeology 1953 0.4854586
##  5: Anthropology &\n Archaeology 1954 0.3885602
##  ---
## 1086:           Others 2014 0.6749932
## 1087:           Others 2015 0.6747165
## 1088:           Others 2016 0.6659535
## 1089:           Others 2017 0.6828912
## 1090:           Others 2018 0.6744691
```



Some facts:

- Between 1950 and 2000, the COC of economics averaged at 0.225, way lower than the average for the SSH at 0.361.
- Between 2000 and 2018, the COC of economics has growth by 20.88 points of percentage (from 0.25 to 0.459), which gives an average growth rate of 0.034.
- Since 2000, most of the other social sciences have also become more outward looking. Taken together, their COC went from 0.406 to 0.472, which corresponds to a growth rate of 0.008.
- Management overtook economics as the least outward looking discipline in 2011. Its COC in 1950, 0.662; in 2000, it was 0.369; and in 2018, it stayed at 0.368.
- The ratio of the COC of economics to the COC of all social sciences in 2000 was 0.615 while, in 2018, it had grown to 0.971.

#### 4.1.2 Weighed citations

Angrist et al. (2020) weigh citations by the ‘importance’ in the discipline of the journal from which they originate. In this subsection, we construct similar weights – customized and normalized impact factors – and then reproduce the analysis from the previous subsection with weighed citations.

We create a customized impact factor with two properties:

1. Only citations from inside of the journal’s discipline count for the impact of this journal.
2. Our impact factor looks back 5 years instead of only 2 years for the standard impact factor. We take a longer window because the temporality of citations in the SSH is longer than in the other sciences.

The combination of these two properties implies that, for journal  $j$ , our impact factor in year  $y$  is equal to:

$$IF_{j,y} = \frac{\#(d_y \rightarrow j_{y-5}) + \dots + \#(d_y \rightarrow j_{y-1})}{\#(j_{y-5}) + \dots + \#(j_{y-1})},$$

where  $d_y \rightarrow j_{y-i}$  is a citation from an article published in year  $y$  in a journal in  $j$ ’s discipline which is directed to an article published in  $j$  in the previous year  $y - i$ . We put a threshold of 25 articles published during the 5 year window to get rid of weird cases (e.g. a journal with only one article, but substantially cited). Under 25 articles published in the previous five years, a journal is not considered.

Finally, we get an annual weight for each journal by normalizing the impact factors (they sum to 1 in each discipline for each year).

We create a table containing, for every journal, the number of articles it cited every year in every journals.

```
dt_citing_cited <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))
discipline_info <- readRDS(paste0(path_to_project_data, "discipline_info.rds"))

setkey(dt_citing_cited,Year, citing_discipline,
       citing_code_revue, cited_discipline, cited_code_revue)
dt_citing_cited_journals <- dt_citing_cited[,list(N = .N), by= key(dt_citing_cited)]
dt_citing_cited_journals <- merge(dt_citing_cited_journals,
                                discipline_info[,list(Code_Discipline,
                                                       Discipline_citing =
                                                       discipline,
                                                       EGrande_Discipline)],
                                by.x = "citing_discipline",
                                by.y = "Code_Discipline")
dt_citing_cited_journals <- merge(dt_citing_cited_journals,
                                discipline_info[,list(Code_Discipline,
```

```

                                Discipline_cited =
                                discipline)],
                                by.x = "cited_discipline",
                                by.y = "Code_Discipline")
dt_citing_cited_journals <- dt_citing_cited_journals[between(Year, min_period,
                                                                max_period)]
setkey(dt_citing_cited_journals, Year, citing_discipline, N)
saveRDS(dt_citing_cited_journals, file=
        paste0(path_to_project_data, "dt_citing_cited_journals.rds"))
rm (dt_citing_cited_journals); rm(dt_citing_cited)

```

We then created the journal weight by the procedure described above.

```

# The standard data with all the citation links at the level of articles
# that involve the SSH:
dt_citing_cited <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))
dt_citing_cited <- adding_disc_name_to_dt_citing_cited(dt_citing_cited)
setnames(dt_citing_cited, c("Year", "cited_year"), c("Year_citing", "Year_cited"))
setkey(dt_citing_cited, Year_citing, citing_code_revue,
        Discipline_citing, Discipline_cited)

# Extracting all the journal_ids
all_j_codes <- c(dt_citing_cited[,unique(citing_code_revue) ],
                dt_citing_cited[,unique(cited_code_revue)]) %>%
  unique() %>% sort()

# Loading a table with the number of articles per year-journal:
nb_articles <- fread(paste0(path_to_project_data,
                            "nb_articles_per_year_and_journal.csv"))
setnames(nb_articles, "Annee_Bibliographique", "Year")
nb_articles <- nb_articles[between(Year, min_period, max_period) &
                          Code_Revue %in% all_j_codes]

# Some parameters
window_length =5

journal_impact_per_window <- data.table(
  citing_Year = numeric(), discipline = character(),
  journal_ID = numeric(), Impact = numeric(), impact_weight = numeric())

for( i in c(min_period:max_period))
{
  # This is to make sure we don't count the impact of a journal
  # that doesn't exist anymore (since we are looking 5 years back)
  code_revue_in_year <- unique(dt_citing_cited[Year_citing == i]$citing_code_revue)
  # Working with a smaller object for the period: year where citations occur
  dt_citing_cited_for_w <- dt_citing_cited[Year_citing == i]

  # Counting number of articles published by every journals
  # in a given period and removing the journals with too few articles.
  nb_art <- nb_articles[Year %in% (i-window_length):(i-1) &
                        Code_Revue %in% code_revue_in_year,

```

```

        list(nb_articles = sum(nb_art)),
        by = Code_Revue][nb_articles > 5*window_length]
# Counting nb refs internal to the discipline to each journal
impact <- dt_citing_cited_for_w[Year_cited %in% (i-window_length):(i-1) &
                                Discipline_citing == Discipline_cited,
                                list(nb_ref = .N),
                                by = .(Discipline_citing, cited_code_revue)]
#Merging nb refs and nb articles
impact <- merge(impact, nb_art, by.x = "cited_code_revue",
                by.y = "Code_Revue", all.x = TRUE)
#Calculating impact
impact <- impact[!is.na(nb_articles)] # removing journals with too few articles
impact[, Impact:= nb_ref / nb_articles]

# Calculating weighted impact (we normalize the impact
# so the weights of every discipline sum to one each year)
impact <- unique(impact[,list(code_revue = cited_code_revue,
                              Discipline_citing, Impact)])
impact[, impact_weight := Impact/sum(Impact), by = "Discipline_citing"]

#Adding a Year column and binding to the journal_impact per window table.
impact$citing_Year <- i
journal_impact_per_window <- rbind(
  journal_impact_per_window,
  impact[,list(citing_Year, discipline = Discipline_citing,
              journal_ID = code_revue, Impact, impact_weight )])
# print(paste0(Sys.time(), " Calculation for normalized impact done for year ", i))
}

# top_5_journal_impact <- journal_impact_per_window %>%
#   group_by(citing_Year, discipline) %>%
#   top_n(n = 5, wt = impact_weight) %>%
#   data.table()
#
# top_5_journal_impact$Top_5 <- TRUE
# journal_impact_per_window <- top_5_journal_impact[journal_impact_per_window,
#   on=(citing_Year, discipline, journal_ID, impact_weight)]
# journal_impact_per_window[is.na(Top_5), Top_5 := FALSE]

setkey(journal_impact_per_window,citing_Year,discipline,impact_weight)
# Producing a flag for the 5% top journals by discipline for each year.
perc_top <- .05
journal_impact_per_window[,Top_5perc := c(rep(FALSE, (.N - max(1 , round(.N *perc_top))))),
                          rep(TRUE, max(1 , round(.N *perc_top)))
                          )
                          , by = .(citing_Year,discipline) ]

save(journal_impact_per_window,
     file=paste0(path_to_project_data, "journal_impact_per_window.RData"))
rm(dt_citing_cited); rm(journal_impact_per_window);
#rm(top_5_journal_impact); rm(dt_citing_cited_for_w); rm(nb_articles)

```

Now, the same graphics as before, but with weighted citations:

```

dt_citing_cited_journals <- readRDS(
  paste0(path_to_project_data, "dt_citing_cited_journals.rds"))
load(paste0(path_to_project_data, "journal_impact_per_window.RData"))
setnames(dt_citing_cited_journals, "Year", "citing_Year")
dt_plot_weighted <- dt_citing_cited_journals[!is.na(Discipline_cited),
  list(citing_code_revue, Discipline_citing,
        Discipline_cited,
        citing_Year, N )][,list(Nb_cited =sum(N)),
  by = list(citing_code_revue,
            Discipline_citing,
            Discipline_cited,
            citing_Year) ]

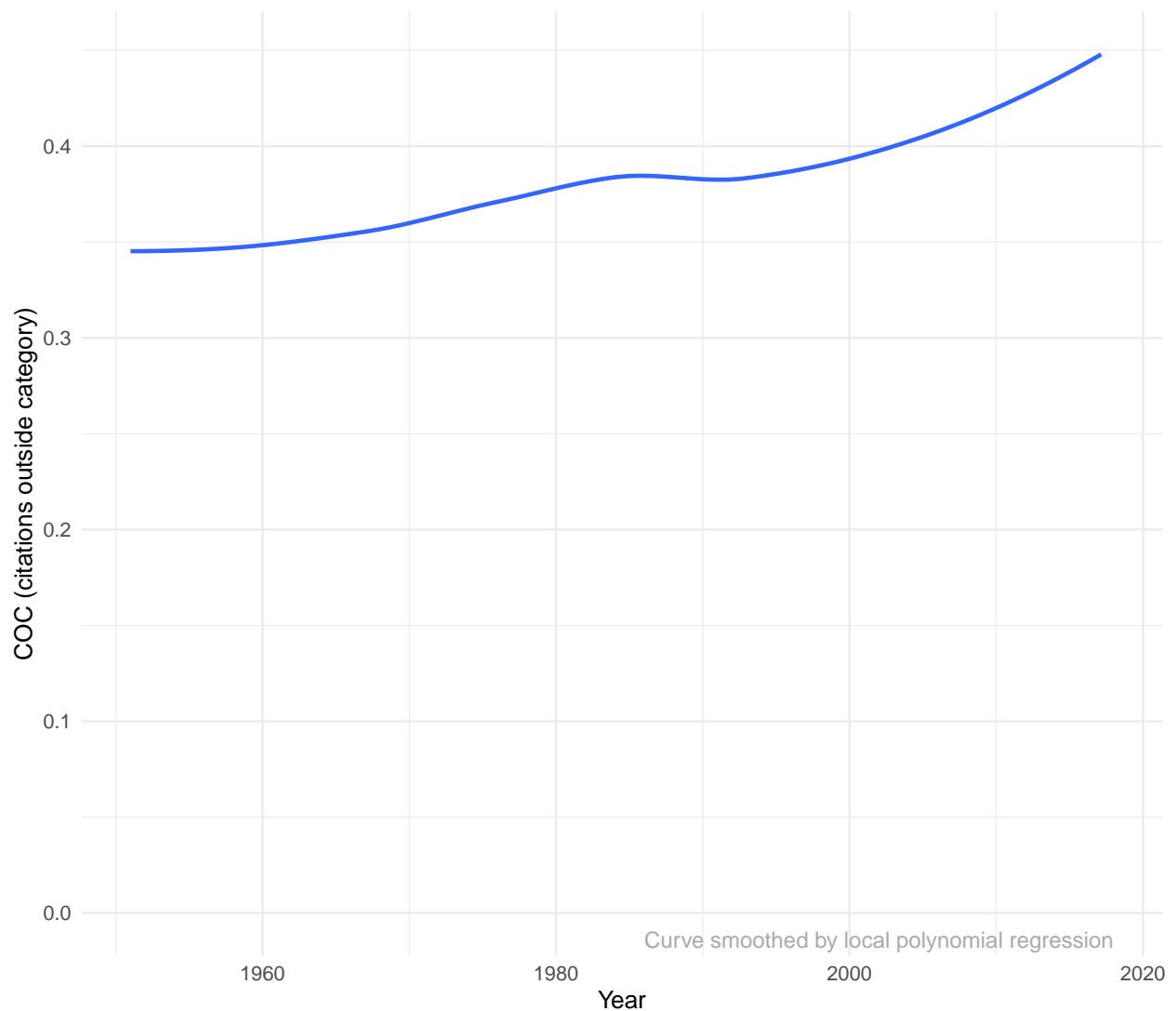
# Getting the weights in there.
dt_plot_weighted <- merge(dt_plot_weighted,
  journal_impact_per_window[,list(discipline,journal_ID,
        citing_Year,
        Journal_weight_in_disc =
        impact_weight )],
  by.x = c("Discipline_citing", "citing_code_revue", "citing_Year"),
  by.y = c("discipline","journal_ID", "citing_Year"))

## Calculating proportions of extradisciplinary citations
# On total SSH:
interdisc_ssh_weighted <- dt_plot_weighted[
  Discipline_citing %in% core_SSH,
  list(Proportion = .SD[Discipline_citing != Discipline_cited,
    sum(Nb_cited*Journal_weight_in_disc)]/sum(
    Nb_cited*Journal_weight_in_disc)
  ),
  by = citing_Year]
setnames(interdisc_ssh_weighted,"citing_Year","Year"); setkey(interdisc_ssh_weighted,Year)

ggplot(interdisc_ssh_weighted,aes(x=Year,y=Proportion)) +
  geom_smooth(se=F, method="loess")+ ylim(0,max(interdisc_ssh_weighted$Proportion)) +
  theme_minimal() +
  theme(axis.text=element_text(size=10),
        axis.title=element_text(size=12 )) +
  annotate( "text", x=max(interdisc_ssh_weighted$Year), y=-Inf,
          hjust=1,vjust=-0.5,
          label="Curve smoothed by local polynomial regression",
          size=4, color="darkgrey" ) +
  ylab("COC (citations outside category)") ##

## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 1 rows containing missing values (`geom_smooth()`).

```



```

# ggtitle("Extradisciplinary citations by the main social sciences")

#Proportion of extradisciplinary citations per discipline:
dt_plot_weighted <- dt_plot_weighted[
  , list(Proportion = .SD[Discipline_citing != Discipline_cited,
    sum(Nb_cited*Journal_weight_in_disc)]/sum(
      Nb_cited*Journal_weight_in_disc)
  ), by = .(citing_Year,Discipline_citing)]

# Finding the order at the end to order in the legend
order_disc <- unique(dt_plot_weighted[
  , Prop := predict(loess(Proportion~citing_Year),
    citing_Year),
  by = Discipline_citing][citing_Year == 2016,][
  order(-Prop),Discipline_citing])
dt_plot_weighted$Discipline_citing <- factor(dt_plot_weighted$Discipline_citing,
  levels = order_disc)

```

```

# Finding the order at the end to order in the legend
setnames(dt_plot_weighted, "citing_Year", "Year")
order_disc <- dt_plot_weighted[Year== 2016][order(-Proportion),Discipline_citing]
dt_plot_weighted$Discipline_citing <- factor(dt_plot_weighted$Discipline_citing,
                                             levels = order_disc)
dt_plot_weighted <- dt_plot_weighted[!is.na(Discipline_citing)]

ymin= min(dt_plot_weighted$Proportion); ymax =max(dt_plot_weighted$Proportion)

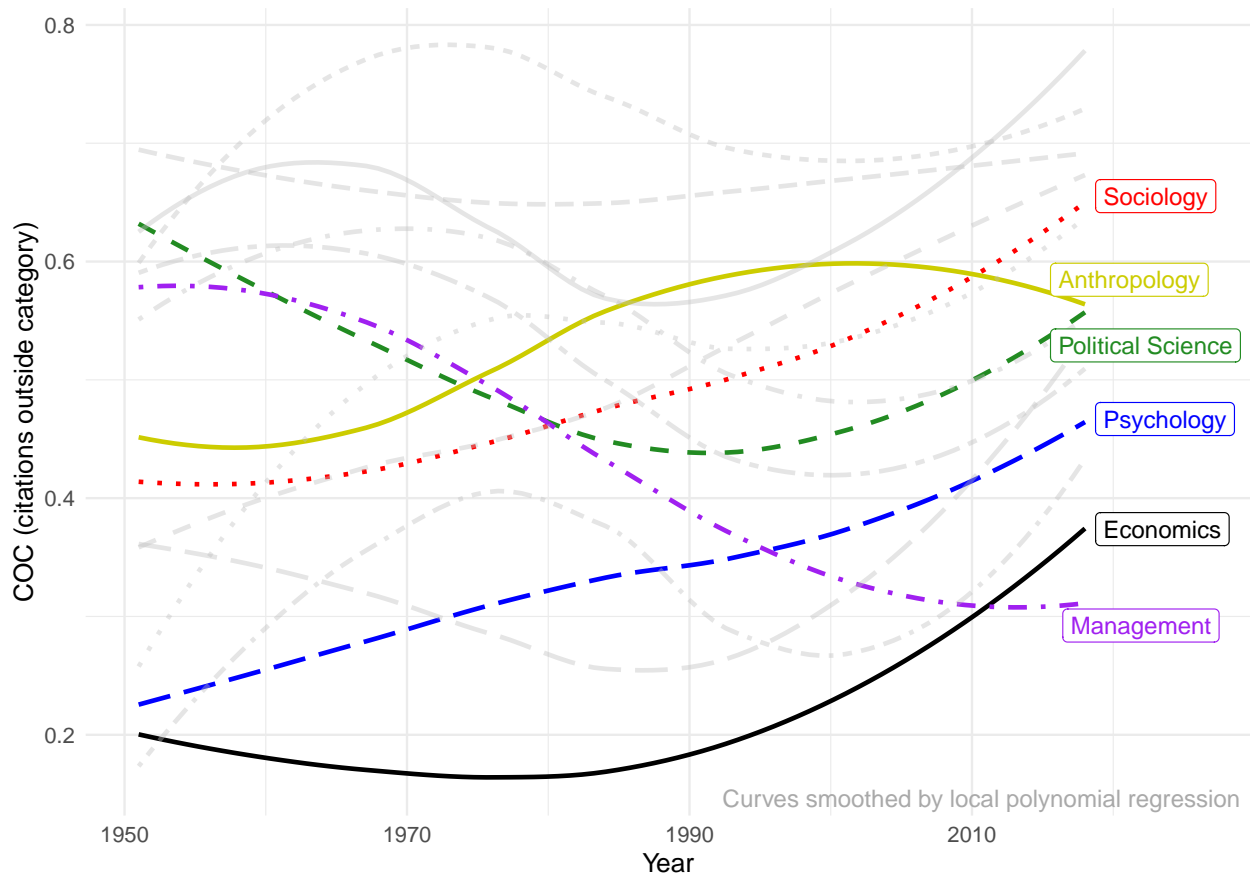
#Abbreviations for the labels at the end of the line
dt_plot_weighted[Discipline_citing == "Psychology", Disc_abbreviation := "Psychology"]
dt_plot_weighted[Discipline_citing == "Economics", Disc_abbreviation := "Economics"]
dt_plot_weighted[Discipline_citing == "Political Science &\nPublic Administration",
                 Disc_abbreviation := "Political Science"]
dt_plot_weighted[Discipline_citing == "Sociology", Disc_abbreviation := "Sociology"]
dt_plot_weighted[Discipline_citing == "Management", Disc_abbreviation := "Management"]
dt_plot_weighted[Discipline_citing == "Anthropology &\n Archaeology",
                 Disc_abbreviation := "Anthropology"]
dt_plot_weighted$Discipline_for_plot <- dt_plot_weighted$Discipline_citing
dt_plot_weighted[!Discipline_citing %in% core_SSH, Discipline_citing := "Others"]

#Plotting
dt_plot_weighted[Discipline_citing %in% core_SSH] %>%
  mutate(label = if_else(Year == max(Year),
                        as.character(Disc_abbreviation), NA_character_)) %>%
  ggplot(mapping = aes(x=Year, y=Prop, col = Discipline_citing)) +
  geom_smooth(se=F,method = 'loess',
            aes(x=Year, y=Prop,col = Discipline_citing, linetype = Discipline_citing),
            show.legend=FALSE) +
  scale_color_manual(name = "Disciplines", breaks = core_SSH, values = colors) +
  scale_linetype_manual(name = "Disciplines", breaks = core_SSH, values = linetypes) +
  # guides(linetype=guide_legend(keywidth = 3, keyheight = 1,nrow=2,byrow=TRUE),
  #        color=guide_legend(keywidth = 3, keyheight = 1,nrow=2,byrow=TRUE)) +
  new_scale_linetype() + # geoms below can use another linetype scale!
  geom_smooth(dt_plot_weighted[Discipline_citing == "Others" &
                               Discipline_for_plot != "Arts"],
            mapping = aes(x=Year, y = Prop#, group = Discipline_for_plot
                          , linetype = Discipline_for_plot),
            color = alpha("gray",.4),
            se=F, method = 'loess') +
  scale_linetype_discrete(name="Other Disciplines") + theme_minimal() +
  theme(legend.position="bottom", legend.box = "vertical",
        axis.text=element_text(size=10), axis.title=element_text(size=12)) +
  ylab("COC (citations outside category)") +
  xlim(c(NA, 2026)) +
  annotate("text",x=max(dt_plot_weighted$Year),y=-Inf,
         hjust=.7,vjust=-0.5,
         label="Curves smoothed by local polynomial regression",
         size=4,color="darkgrey" ) +
  guides(linetype=guide_legend(keywidth = 3, keyheight = 1,nrow=4,byrow=TRUE)) +
  geom_label_repel(aes(label = label), nudge_x = 1,
                 xlim = c(2015, NA), na.rm = TRUE, show.legend = FALSE)

## `geom_smooth()` using formula = 'y ~ x'

```

```
## `geom_smooth()` using formula = 'y ~ x'
```



— Demography	- - - Other Social Sciences	- - - Health
- - - Other Professional Fields	· · · Geography	· · · International Relations
- - - Humanities	- - - Education	· · · Law

```
shades_of_gray <- c("Psychology" = "#404040",
  "Economics" = "black",
  "Sociology" = "#202020",
  "Anthropology &\n Archaeology" = "#A0A0A0",
  "Political Science &\n Public Administration" = "#606060",
  "Management" = "808080",
  "Others" = alpha("gray",.3))

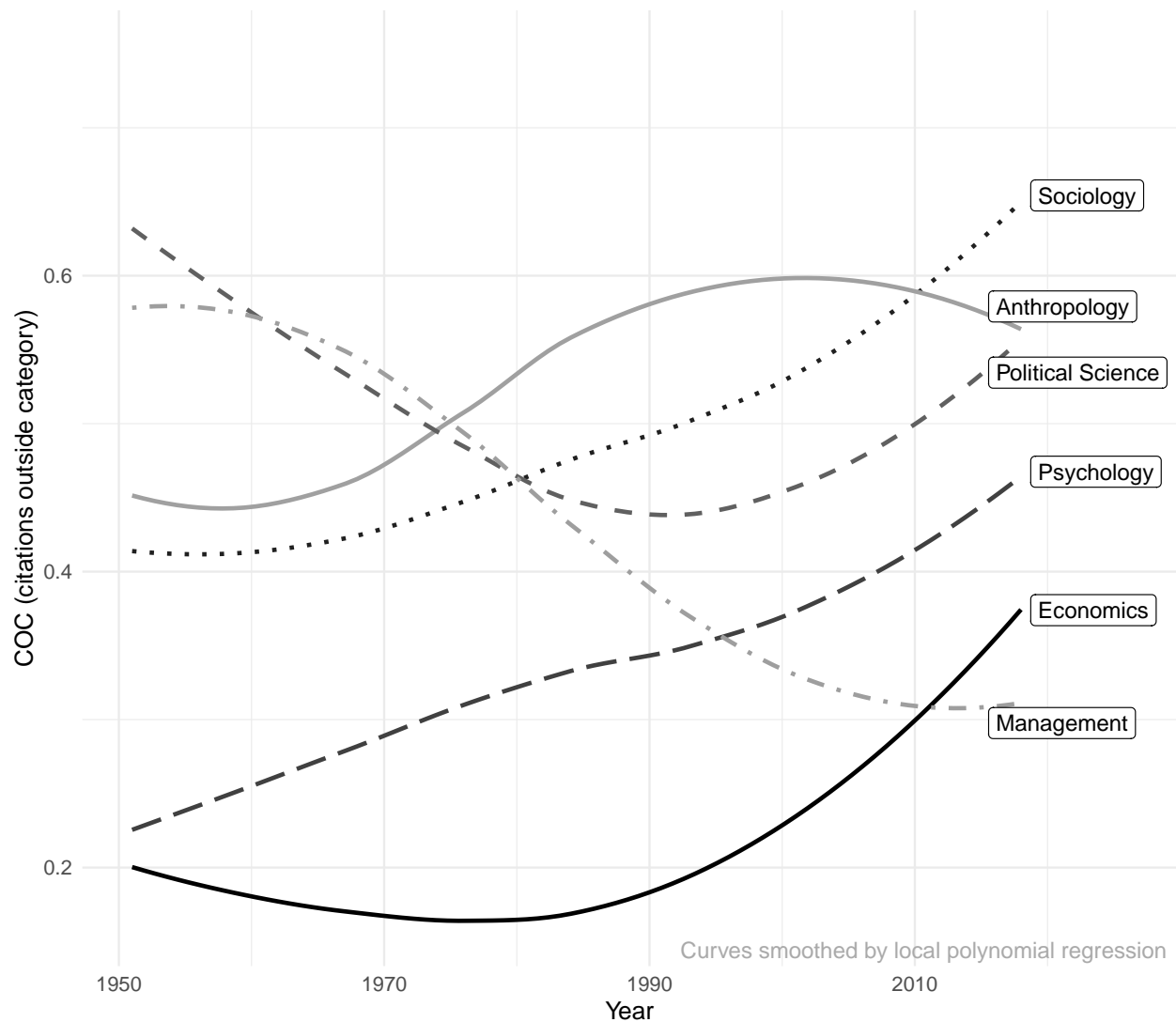
dt_plot_weighted[Discipline_citing %in% core_SSH] %>%
  mutate(label = if_else(Year == max(Year),
    as.character(Disc_abbreviation), NA_character_)) %>%
  ggplot(mapping = aes(x=Year, y=Prop, gr = Discipline_citing)) +
  geom_smooth(se=F,method = 'loess',
    aes(x=Year, y=Prop,col = Discipline_citing, linetype = Discipline_citing),
    show.legend=FALSE) +
  scale_color_manual(name = "Disciplines", breaks = core_SSH, values = shades_of_gray) +
  scale_linetype_manual(name = "Disciplines", breaks = core_SSH, values = linetypes) +
  theme_minimal() +
  # guides(linetype=guide_legend(keywidth = 3, keyheight = 1,nrow=2,byrow=TRUE),
```

```

#       color=guide_legend(keywidth = 3, keyheight = 1,nrow=2,byrow=TRUE)) +
theme(legend.position="bottom", legend.box = "vertical",
      axis.text=element_text(size=10), axis.title=element_text(size=12)) +
ylab("COC (citations outside category)") +
xlim(c(NA, 2026)) +
ylim(c(NA, 0.75)) +
annotate( "text", x=max(dt_plot_weighted$Year), y=-Inf,
          hjust=.7, vjust=-0.5,
          label="Curves smoothed by local polynomial regression",
          size=4,color="darkgrey" ) +
geom_label_repel(aes(label = label), nudge_x = 1, xlim = c(2015, NA),
                na.rm = TRUE, show.legend = FALSE)

```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
rm(dt_citing_cited_journals)
```

```
dt_plot_weighted[,list(Discipline_citing, Year, Prop )]
```

```
##           Discipline_citing Year      Prop
```



```
## 1: Anthropology &\n Archaeology 1962 0.4330935
## 2: Anthropology &\n Archaeology 1964 0.4360734
## 3: Anthropology &\n Archaeology 1965 0.4383324
## 4: Anthropology &\n Archaeology 1966 0.4409796
## 5: Anthropology &\n Archaeology 1967 0.4439214
## ---
## 1071: Sociology 1963 0.4125659
## 1072: Sociology 1964 0.4137102
## 1073: Sociology 1965 0.4150475
## 1074: Sociology 1966 0.4165554
## 1075: Sociology 1967 0.4182115
```

Looking at the same facts:

- Between 1950 and 2000, the COC of economics averaged at 0.181, way lower than the average for the SSH at 0.367.
- Between 2000 and 2018, the COC of economics has growth by 14.54 points of percentage (from 0.229 to 0.374), which gives an average growth rate of 0.027.
- Since 2000, most of the other social sciences have also become more outward looking. Taken together, their COC went from 0.384 to 0.448, which corresponds to a growth rate of 0.009.
- Management overtook economics as the least outward looking discipline in 2011. Its COC in 1951, 0.565; in 2000, it was 0.331; and in 2018, it stayed at 0.315.
- The ratio of the COC of economics to the COC of all social sciences in 2000 was 0.596 while, in 2018, it had grown to 0.835.

We can generalize that the same trends toward increased interdisciplinarity and a catching up with the other social sciences is visible even when we used journal weights, but that these trends are less pronounced.

## 4.2 Internal hierarchy (and reproducing Fourcade et al. 2015)

The question now is whether the above results for whole disciplines (with the unweighted COC) hold for top journals in these disciplines.

Focusing on economics, our first way to answer this question is to contrast the evolution of extradisciplinary citations for all of economics to the evolution of this measure for the selection of top journals made by Fourcade et al. (2015). This selection is:

1. Quarterly Journal of Economics
2. Journal of Political Economy
3. American Economic Review
4. Econometrica
5. Review of Economic Studies

Since we will be working with data similar to Fourcade et al. (2015), we take the opportunity to reproduce their Figure 3 with our data:

```
last_year = max_period

#We load the big object with all the
dt_citing_cited <- readRDS(
  paste0(path_to_project_data, "dt_citing_cited_all.rds"))

dt_citing_cited <- dt_citing_cited[!is.na(cited_discipline) &
```

```
Year %in% min_period:last_year]
```

```
# We get the journals:
```

```
ESH <- dbConnect(MySQL(), user=usr, password=pswd, dbname='OST_Expanded_SciHum',  
                host='127.0.0.1')
```

```
main_j <- dbGetQuery(ESH,  
  paste0('SELECT DISTINCT code_revue FROM OST_Expanded_SciHum.Revues  
        WHERE Revue like "Quarterly Journal of Economics%" or  
              Revue like "Journal of Political Economy%" or  
              Revue like "American Economic Review%" or  
              Revue like "Econometrica%" or  
              Revue like "Review of Economic Studies%"'))
```

```
art_journal <- dt_citing_cited[citing_code_revue %in% main_j$code_revue]
```

```
setkey(art_journal,cited_discipline); setkey(discipline_info, Code_Discipline)
```

```
art_journal <- merge(art_journal,  
  discipline_info[,list(Code_Discipline,discipline)],  
  by.x="cited_discipline", by.y="Code_Discipline",all.x=T)
```

```
setkey(art_journal, Year, discipline)
```

```
citing_cited_selected_j <- art_journal[, list(Citations =.N), by=key(art_journal)]
```

```
citing_cited_selected_j <- citing_cited_selected_j[  
  ,list(Discipline_cited = discipline, Proportion=Citations/sum(Citations)),  
  by=Year]
```

```
# Finding out the most important disciplines:
```

```
the_disc = "Economics"; the_varia <- "All others"
```

```
imp_disc <- citing_cited_selected_j[Discipline_cited != the_disc &  
  between(Year, firstYearImpDisc, last_year),  
  list(tot_prop_disc = sum(Proportion)),  
  by=Discipline_cited][order(-tot_prop_disc)][  
  1:n_disc, Discipline_cited]
```

```
dt_p <- citing_cited_selected_j[Discipline_cited %in% imp_disc]
```

```
dt_p <- rbind(dt_p,citing_cited_selected_j[!Discipline_cited %in% c(imp_disc,the_disc),  
  list(Discipline_cited=the_varia,  
        Proportion= sum(Proportion)),by=Year])
```

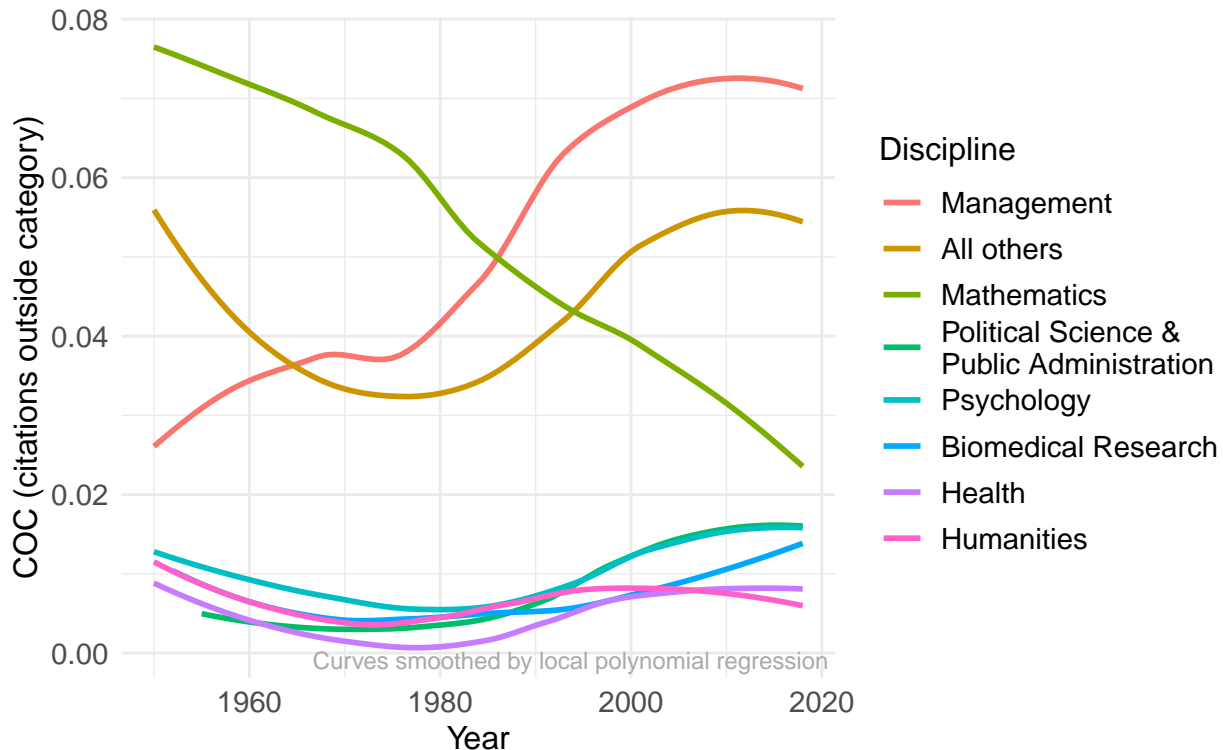
```
order_disc <- dt_p[, Prop := predict(loess(Proportion~Year), Year),  
  by = Discipline_cited][Year == last_year,][  
  order(-Prop),Discipline_cited]
```

```
setkey(dt_p,Year,Discipline_cited)
```

```
dt_p$Discipline_cited <- factor(dt_p$Discipline_cited, levels = order_disc)

ggplot(dt_p, aes(x=Year, y=Proportion, col = Discipline_cited)) +
  geom_smooth(se=F,method = 'loess') + theme_minimal() +
  labs(col = "Discipline") + #ylim(0, 0.12) +
  ylab("COC (citations outside category)") +
  #ggtitle("Extradisciplinary citations from five top journals in economics") +
  theme(axis.text=element_text(size=11),
        axis.title=element_text(size=12),
        legend.text=element_text(size=11),
        legend.title=element_text(size=12)) +
  annotate("text",x=max(dt_p$Year),y=-Inf,hjust=.95,vjust=-0.5,
        label="Curves smoothed by local polynomial regression",
        size=3,color="darkgrey" )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
ggsave( paste0("figures/", "fig15b_our_version.png"), width = 16, height = 7, bg = "white",
        scale = .5)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
rm(dt_citing_cited, main_j, art_journal, dt_p)
```

Note that this graph is used late in our article (section 6), when we compare our results to the existing literature.

Fourcade et al. (2015, p. 102) say: “Our analysis of five top economics journals shows that between 19 and 25 percent of citations are outside the discipline, a fairly stable pattern since the end of World War II.” This claim clashes with our result above that economics has been citing more other disciplines recently. Do we have Fourcade’s results if we focus on the top 5?

```

dt_interdisc_econ <- citing_cited_selected_j[Discipline_cited=="Economics" &
  Year<=last_year,
  list(Year, Corpus="Top 5 Journals",
  Proportion= 1-Proportion)]

temp <- citing_cited_disc[!is.na(cited_discipline) & citing_discipline=="119 ]

temp <- temp[,list(Discipline_cited, Proportion= Citations/sum(Citations)),by=Year]

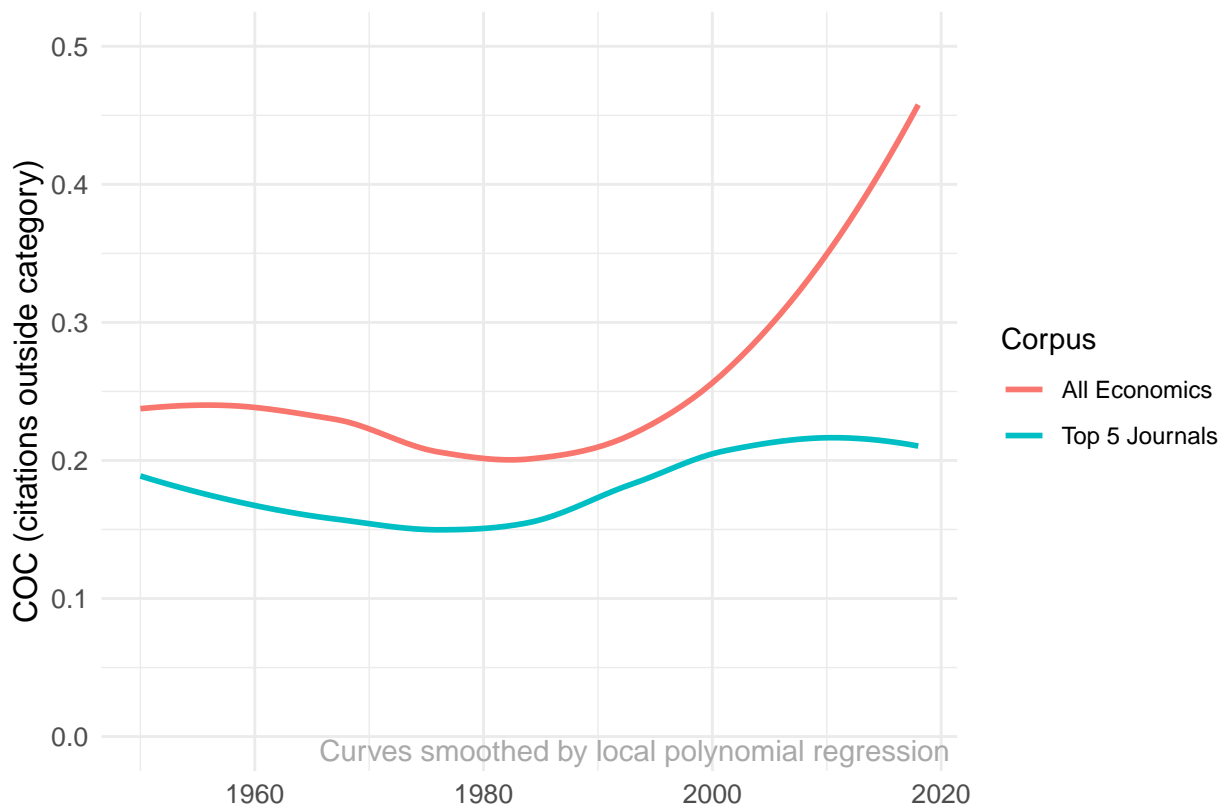
temp <- temp[Discipline_cited=="Economics", list(Year,
  Corpus="All Economics",
  Proportion= 1-Proportion)]

dt_interdisc_econ <- rbind(dt_interdisc_econ,temp) ; rm(temp)

ggplot(dt_interdisc_econ[Year<=last_year],aes(x= Year,y= Proportion, color = Corpus)) +
  geom_smooth(se=F,method="loess") + theme_minimal() +
  xlab("")+ ylab("COC (citations outside category)") + ylim(0,.5) +
  theme(axis.text=element_text(size=10),
  axis.title=element_text(size=12#,face="bold
  )) +
  annotate( "text",x=max(dt_interdisc_econ$Year),y=-Inf,hjust=.95,vjust=-0.5,
  label="Curves smoothed by local polynomial regression",
  size=4,color="darkgrey" ) #+

```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
# ggtitle("Extradisciplinary citations of Economics")
```

```
ggsave( paste0("figures/", "fig4.png"), width = 16, height = 9, bg = "white", scale=.5)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Yes, that looks about right for the top 5 journals. In fact, the proportion only varies between 0.11 and 0.25.

The recent split between the citation pattern of the top journals and of the rest of economics is quite interesting. It is the first time since the 1950s that the split is that important.

```
dt_top5_vs_all_econ <- spread(dt_interdisc_econ,"Corpus","Proportion") %>% data.table()
```

```
dt_top5_vs_all_econ[,Diff_in_interdisc:= `All Economics`- `Top 5 Journals`]
```

```
dt_top5_vs_all_econ
```

##	Year	All Economics	Top 5 Journals	Diff_in_interdisc
##	1: 1950	0.2044888	0.2301255	-0.025636745
##	2: 1951	0.1800643	0.1391304	0.040933874
##	3: 1952	0.2875989	0.1813725	0.106226396
##	4: 1953	0.2072072	0.1882353	0.018971913
##	5: 1954	0.1825558	0.1115108	0.071044990
##	6: 1955	0.2607945	0.2511628	0.009631683
##	7: 1956	0.2741703	0.1282051	0.145965146
##	8: 1957	0.3029240	0.1699346	0.132989336
##	9: 1958	0.2981818	0.2131148	0.085067064
##	10: 1959	0.2500000	0.1680912	0.081908832
##	11: 1960	0.2478109	0.2000000	0.047810858
##	12: 1961	0.2446881	0.1713147	0.073373402
##	13: 1962	0.1990391	0.1272727	0.071766394
##	14: 1963	0.2099553	0.1494083	0.060547045
##	15: 1964	0.3063357	0.2208333	0.085502395
##	16: 1965	0.2575156	0.1340942	0.123421447
##	17: 1966	0.2315186	0.1768617	0.054656923
##	18: 1967	0.2504939	0.1219839	0.128509962
##	19: 1968	0.2649481	0.1886792	0.076268839
##	20: 1969	0.1962884	0.1417134	0.054574995
##	21: 1970	0.2310434	0.1478979	0.083145460
##	22: 1971	0.1949721	0.1455064	0.049465648
##	23: 1972	0.1819584	0.1620654	0.019892926
##	24: 1973	0.2127485	0.1634820	0.049266505
##	25: 1974	0.2021337	0.1459040	0.056229673
##	26: 1975	0.2179050	0.1435185	0.074386527
##	27: 1976	0.2232246	0.1538788	0.069345869
##	28: 1977	0.1859859	0.1565068	0.029479045
##	29: 1978	0.2025298	0.1781331	0.024396691
##	30: 1979	0.1899257	0.1212023	0.068723407
##	31: 1980	0.2140450	0.1626355	0.051409445
##	32: 1981	0.1920758	0.1342303	0.057845509
##	33: 1982	0.2028964	0.1458869	0.057009548
##	34: 1983	0.2057979	0.1619327	0.043865295
##	35: 1984	0.2028797	0.1541551	0.048724659
##	36: 1985	0.2377957	0.1434263	0.094369411
##	37: 1986	0.1960577	0.1850105	0.011047256

## 38: 1987	0.1979173	0.1423002	0.055617145
## 39: 1988	0.2012011	0.1517011	0.049500052
## 40: 1989	0.1990295	0.1671827	0.031846882
## 41: 1990	0.2062830	0.1833417	0.022941379
## 42: 1991	0.2087647	0.1763326	0.032432075
## 43: 1992	0.2361904	0.1525632	0.083627275
## 44: 1993	0.2092474	0.1690418	0.040205630
## 45: 1994	0.2138717	0.1808374	0.033034343
## 46: 1995	0.2282595	0.1937046	0.034554856
## 47: 1996	0.2405429	0.2198617	0.020681260
## 48: 1997	0.2474446	0.1918542	0.055590331
## 49: 1998	0.2461783	0.2422283	0.003949964
## 50: 1999	0.2524682	0.2026873	0.049780887
## 51: 2000	0.2497639	0.1779680	0.071795871
## 52: 2001	0.2490039	0.2174085	0.031595372
## 53: 2002	0.2698660	0.1771526	0.092713435
## 54: 2003	0.2641867	0.2530817	0.011105019
## 55: 2004	0.2774350	0.2390953	0.038339728
## 56: 2005	0.2839393	0.2421432	0.041796078
## 57: 2006	0.3006859	0.2284707	0.072215271
## 58: 2007	0.3169515	0.2082345	0.108716992
## 59: 2008	0.3336494	0.2469560	0.086693459
## 60: 2009	0.3526005	0.2122495	0.140351001
## 61: 2010	0.3437837	0.2085915	0.135192197
## 62: 2011	0.3618362	0.2176012	0.144235028
## 63: 2012	0.3868528	0.2010505	0.185802273
## 64: 2013	0.3970374	0.2143122	0.182725200
## 65: 2014	0.3979562	0.1855465	0.212409712
## 66: 2015	0.4086686	0.2292096	0.179459090
## 67: 2016	0.4243225	0.2191483	0.205174158
## 68: 2017	0.4399384	0.2030791	0.236859301
## 69: 2018	0.4585662	0.2054231	0.253143132
##	Year	All Economics Top 5 Journals	Diff_in_interdisc

From 1950 to 2000, the average difference between the two was 0.057. At the end of the period, it had grown to 0.253.

Our second way to study the interaction between internal hierarchy and interdisciplinarity is to use our impact factor to isolate, for each year, the journals in the top 5% of each discipline and to contrast the COC of these journals to the COC of the rest of the discipline.

```
dt_citing_cited_all <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))
dt_citing_cited_all <- adding_disc_name_to_dt_citing_cited(dt_citing_cited_all)
setkey(dt_citing_cited_all, Year, citing_code_revue, Discipline_citing, Discipline_cited)
dt_citing_cited_journals <- readRDS(
  paste0(path_to_project_data, "dt_citing_cited_journals.rds"))
load(paste0(path_to_project_data, "journal_impact_per_window.RData"))
core_SSH <- c("Psychology", "Economics",
              "Political Science &\nPublic Administration",
              "Sociology", "Anthropology &\n Archaeology",
              "Management")

# Finding the COC for each set of journals (top 5% vs rest) in all years
setnames(dt_citing_cited_journals, "Year", "citing_Year")
setkey(dt_citing_cited_journals, citing_Year, citing_code_revue)
```

```

setkey(journal_impact_per_window,citing_Year, journal_ID)

dt_citing_cited_journals <- dt_citing_cited_journals[Discipline_citing %in% core_SSH]
top_bottom <- merge(dt_citing_cited_journals[,list(citing_Year,Discipline_citing,
citing_discipline, citing_code_revue,
Discipline_cited,cited_discipline,N)],
journal_impact_per_window[,list(citing_Year,journal_ID,
impact_weight,Top_5perc)],
by.x= c("citing_Year","citing_code_revue"),
by.y= c("citing_Year","journal_ID"), all.x = TRUE)
top_bottom[is.na(Top_5perc), `:=`(Top_5perc = FALSE, impact_weight = 0)]
top_bottom <- top_bottom[,list(Total_out_disc= .SD[
Discipline_citing!=Discipline_cited,sum(N)]
,Total_disc= sum(N)
),by=(citing_Year,Discipline_citing,Top_5perc)]

top_bottom[,COC:=Total_out_disc/Total_disc]

# Preparing for plot
top_bottom[Top_5perc ==TRUE, Journals := "Top 5% of journals"]
top_bottom[Top_5perc ==FALSE, Journals := "All others"]
top_bottom$Top_5perc <- NULL

top_bottom[Journals == "All others"]$COC <- -top_bottom[Journals == "All others"]$COC
top_bottom$Journals <- factor(top_bottom$Journals,
levels = c("Top 5% of journals", "All others"))
setkey(top_bottom, citing_Year, Discipline_citing)

top_bottom[, `:=`(
low_bound =prop.test(x= c(.SD[Journals == "Top 5% of journals",
Total_out_disc],.SD[Journals == "All others",
Total_out_disc]),
n = c(.SD[Journals == "Top 5% of journals",
Total_disc],.SD[Journals == "All others",
Total_disc]),
alternative = "two.sided",correct=FALSE)$conf.int[1],
high_bound =prop.test(x= c(.SD[Journals == "Top 5% of journals",
Total_out_disc],.SD[Journals == "All others",
Total_out_disc]),
n = c(.SD[Journals == "Top 5% of journals",
Total_disc],.SD[Journals == "All others",
Total_disc]),
alternative = "two.sided",correct=FALSE)$conf.int[2],
balance = sum(COC)
), by= .(Discipline_citing,citing_Year)]

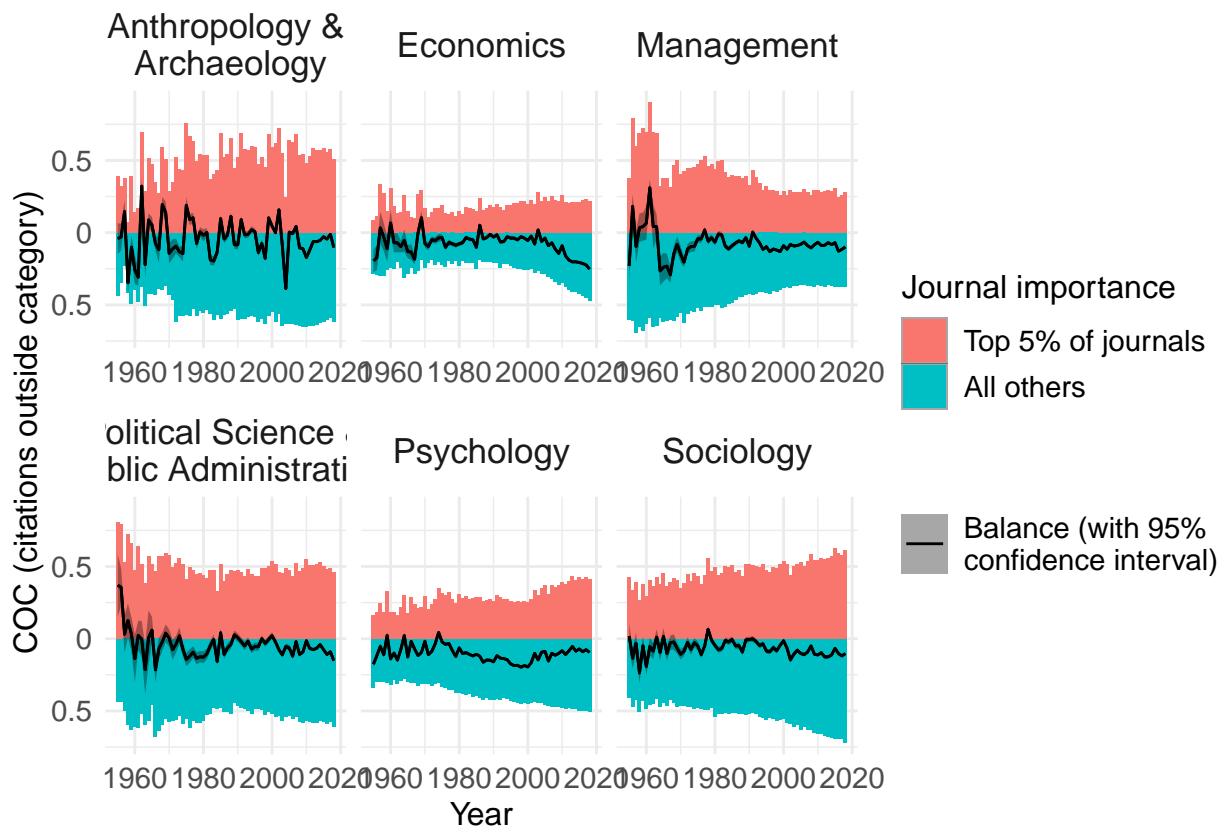
#Plotting
min_y = 1955 # if we start too early, we get a lot of movement
a <- .35 ; color_interval <- "black"
ggplot(top_bottom[Discipline_citing %in% core_SSH & citing_Year >=min_y],
aes(x=citing_Year, y=COC, fill=Journals)) +

```

```

geom_bar(stat="identity", position="identity") +
facet_wrap(~ Discipline_citing, scales = "free_x") +
scale_fill_discrete(name="Journal importance",
                    labels=c("Top 5% of journals", "All others")) +
geom_line(aes(x=citing_Year, y=balance,
              color = "Balance (with 95%\nconfidence interval)")) +
geom_ribbon(aes(ymin=low_bound, ymax=high_bound),
           fill=color_interval, alpha=a) + theme_minimal() +
# ggtitle("Extradiisciplinary citations proportion for top 5% journals vs bottom 95%") +
xlab("Year") + ylab("COC (citations outside category)") +
scale_y_continuous(breaks = pretty(top_bottom$COC),
                  labels = abs(pretty(top_bottom$COC))) +
scale_color_manual(name = "",
                  values = c("Balance (with 95%\nconfidence interval)" = "black")) +
theme(legend.key = element_rect(colour = "transparent",
                                fill = alpha(color_interval, a)),
      axis.text=element_text(size=11),
      axis.title=element_text(size=12),
      legend.text=element_text(size=11),
      legend.title=element_text(size=12),
      strip.text.x = element_text(size = 13))

```



```

ggsave( paste0("figures/", "fig5.png"), width = 16, height = 9, bg = "white",
        scale = .6)

```

```

rm(dt_citing_cited_journals)

```



```

# We might want to keep the top journals per year

# Loading mapping between journals and disciplines:
# ESH <- dbConnect(MySQL(), user=usr, password=pswd, dbname='OST_Expanded_SciHum',
#                 host='127.0.0.1')
# disc_w_journal <- dbGetQuery(ESH,
# "SELECT Code_Revue as journal_ID, Revue as Journal
# FROM OST_Expanded_SciHum.Revues;") %>% data.table
# disc_w_journal$Journal <- strsplit(disc_w_journal$Journal, "\r") %>% unlist()
#
# top_journals <- merge(journal_impact_per_window[Top_5==TRUE],
# disc_w_journal, by="journal_ID")
#
# top_journals[, list(Discipline=discipline,
# Year= citing_Year, Journal, Proportion=round(impact_weight,3))][
# order(Discipline, Year, -Proportion)]

rm(journal_impact_per_window) ; rm(disc_w_journal); rm(dt_citing_cited_all);

```

The balance does not show exactly clearly how far the top 5% is from the bottom 95% because we do not have the ratio of the two COC. We thus produce another figure with the ratio of COC for our main disciplines:

```

dt_plot_ratio_COC_top_rest <- top_bottom[citing_Year >=min_y, list(
  log_Ratio_COC = log(.SD[Journals=="Top 5% of journals", COC]/
    .SD[Journals=="All others", -COC])),
  by=. (Discipline_citing, citing_Year)]

the_span <- .8
label_plot <- dt_plot_ratio_COC_top_rest[
  , list(y = loess(log_Ratio_COC ~ citing_Year,
    span = the_span, data = .SD) %>%
    predict(max(citing_Year), x= max(citing_Year)),
  by= Discipline_citing]
label_plot[, label:= Discipline_citing]
label_plot[grepl("Anthropology", label), label := "Anthropology"]
label_plot[grepl("Political Science", label), label := "Political Science"]

ggplot(dt_plot_ratio_COC_top_rest, aes(x=citing_Year,
  y=log_Ratio_COC, color = Discipline_citing)) +
  geom_smooth(se=F, method = 'loess', span = the_span,
  aes(x=citing_Year, y = log_Ratio_COC, color = Discipline_citing,
  linetype = Discipline_citing), show.legend=FALSE) +
  scale_color_manual(name = "Disciplines", breaks = core_SSH, values = colors) +
  scale_linetype_manual(name = "Disciplines", breaks = core_SSH,
  values = linetypes) + theme_minimal() +
  # guides(linetype=guide_legend(keywidth = 3, keyheight = 1, nrow=2, byrow=TRUE),
  # color=guide_legend(keywidth = 3, keyheight = 1, nrow=2, byrow=TRUE)) +
  theme(legend.position="bottom", legend.box = "vertical",
  axis.text=element_text(size=12), axis.title=element_text(size=13)) +
  ylab("log of COC ratios") +
  scale_x_continuous("Year", breaks = seq(1960, 2020, 10), limits = c(NA, 2029)) +
  # ylim(c(NA, 1.2)) +

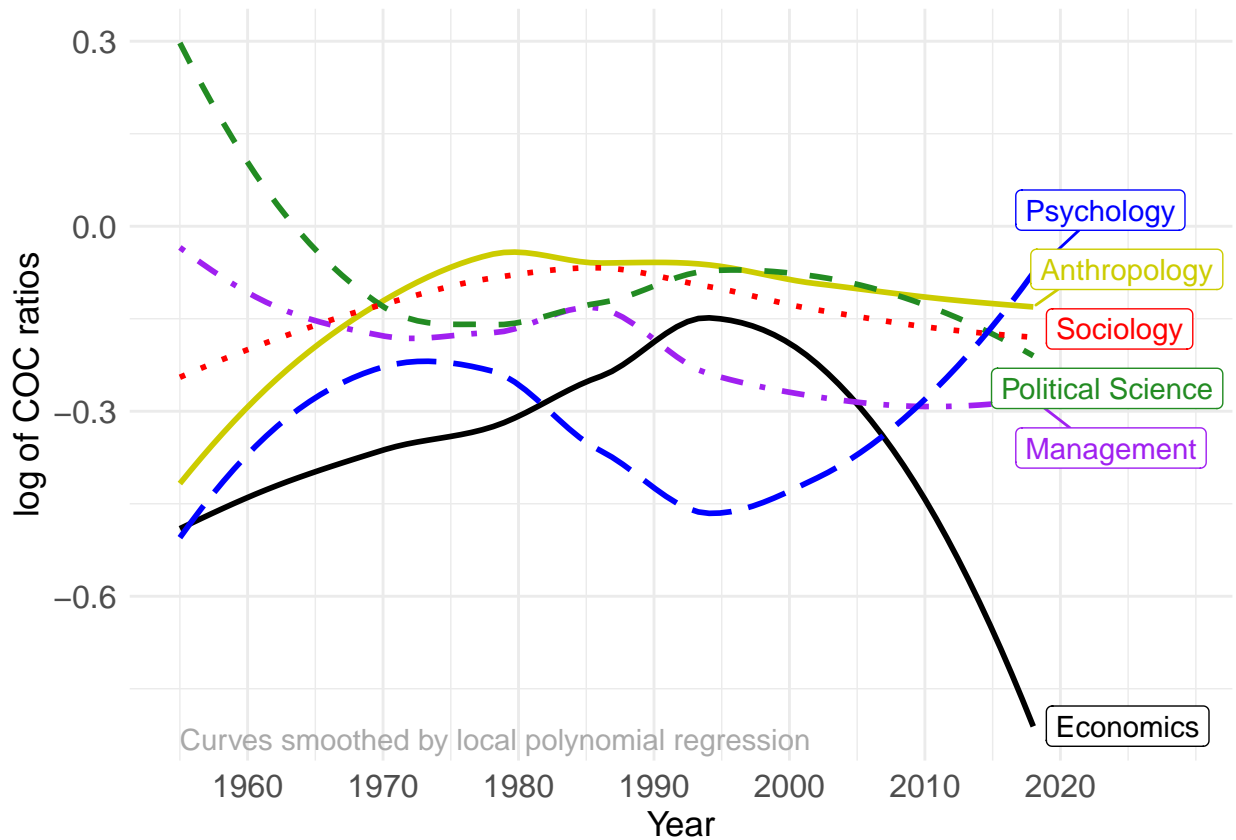
```

```

annotate( "text",x=min(dt_plot_ratio_COC_top_rest$citing_Year),y=-Inf,
          hjust=0,vjust=-0.5,
          label="Curves smoothed by local polynomial regression",
          size=4,color="darkgrey" ) +
geom_label_repel(label_plot, mapping = aes(x= x, y= y, label = label),
                 nudge_x = 1, xlim = c(2016, NA), na.rm = TRUE,
                 show.legend = FALSE)

```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```

ggsave( paste0("figures/", "fig6.png"), width = 16, height = 9, bg = "white",
         scale = .6)

```

```
## `geom_smooth()` using formula = 'y ~ x'
```

### 4.3 Citations toward disciplines

Our next step is to look at citations toward a discipline (CTC), which is a way to measure the extent to which citations to a discipline come from the outside of this discipline.

```

core_SSH <- c("Psychology","Economics",
              "Political Science &\nPublic Administration",
              "Sociology", "Management",
              "Anthropology &\n Archaeology")

dt_citing_cited <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))

```

```

dt_citing_cited <- dt_citing_cited[between(Year, 1950, max_period)]
dt_citing_cited <- merge(dt_citing_cited,
                        discipline_info[,list(Code_Discipline,
                                             Discipline_cited = discipline)],
                        by.x="cited_discipline", by.y="Code_Discipline", all.x=T)
dt_citing_cited <- merge(dt_citing_cited,
                        discipline_info[,list(Code_Discipline,
                                             Discipline_citing = discipline)],
                        by.x="citing_discipline", by.y="Code_Discipline", all.x=T)

CTC_per_y_and_disc <- dt_citing_cited[!is.na(cited_discipline) &
                                     !is.na(citing_discipline)]
CTC_per_y_and_disc <- unique(CTC_per_y_and_disc[,list(Discipline_cited, Citations = .N),
                                                    by= list(Year,Discipline_citing,
                                                            Discipline_cited)])
CTC_per_y_and_disc <- CTC_per_y_and_disc[,list(Discipline_citing,
                                              Proportion = Citations/sum(Citations)),
                                         by= list(Year,Discipline_cited)]

setkey(CTC_per_y_and_disc, Year, Discipline_cited)
# The plot data, focussing on core SSH
dt_plot <- CTC_per_y_and_disc[Discipline_citing==Discipline_cited,
                             list(Year, Discipline = Discipline_cited,
                                   Prop = 1-Proportion)]

# Smoothing y coordinate outside ggplot with loess
dt_plot <- dt_plot[,list(Year, Prop, Proportion = predict(loess(Prop~Year), Year)),
                  by=Discipline]

# Finding the order at the end to order in the legend
order_disc <- dt_plot[Year== 2016][order(-Proportion),Discipline]
dt_plot$Discipline <- factor(dt_plot$Discipline, levels = order_disc)
dt_plot <- dt_plot[!is.na(Discipline)]
ymin= min(dt_plot$Proportion); ymax =max(dt_plot$Proportion)

# Colors for every disciplines. We used Angrist's colors and added purple for Management
colors <- c("Psychology" = "#0000FF",
            "Economics" = "black",
            "Sociology" = "#FF0000",
            "Anthropology &\n Archaeology" = "#CCCC00",
            "Political Science &\nPublic Administration" = "#228B22",
            "Management" = "purple",
            "Others" = alpha("gray",.3))

# Linetypes for every disciplines. We used Angrist's linetype
# and added dotdash for Management
linetypes <- c("Psychology" = "longdash",
              "Economics" = "solid",
              "Sociology" = "dotted",
              "Anthropology &\n Archaeology" = "solid",
              "Political Science &\nPublic Administration" = "dashed",
              "Management" = "dotdash",

```

```

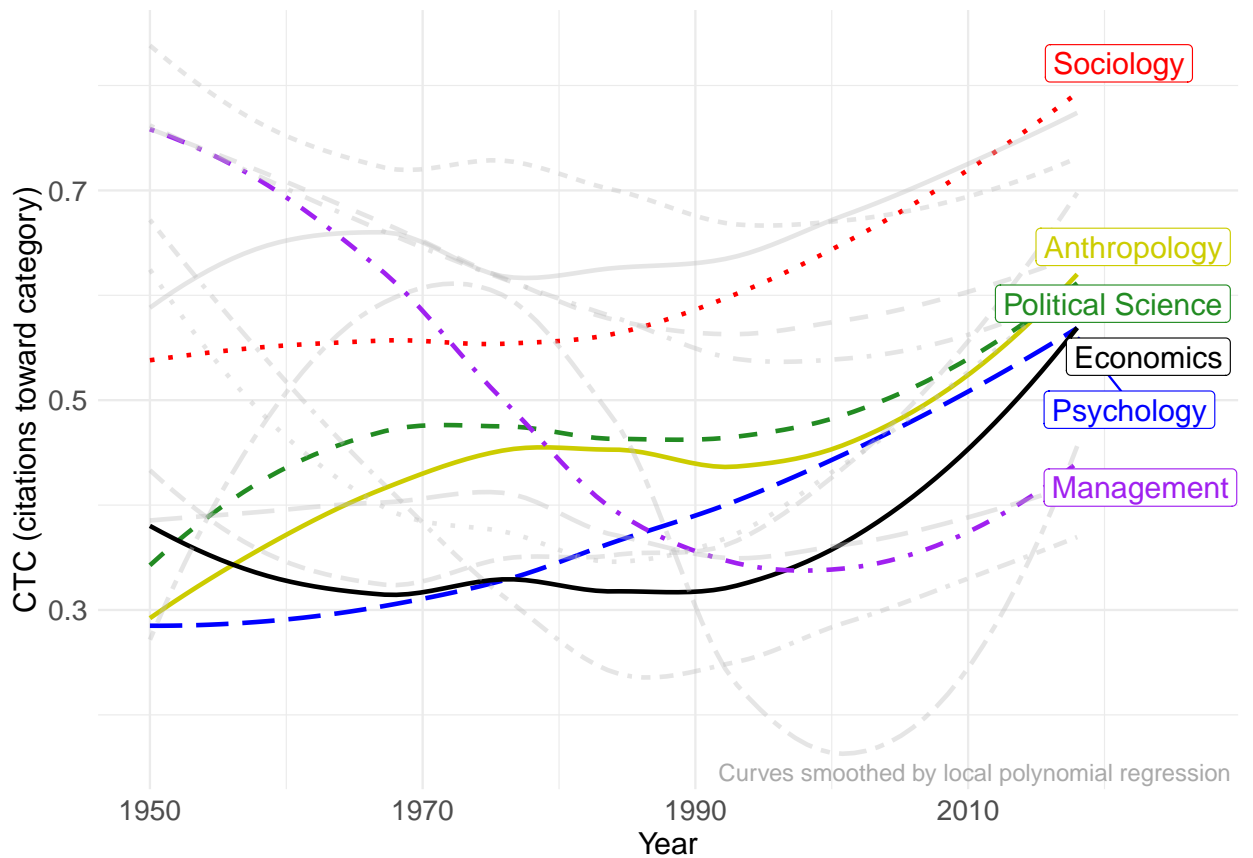
    "Others" = "solid")

#Abbreviations for the labels at the end of the line
dt_plot[Discipline == "Psychology", Disc_abbreviation := "Psychology"]
dt_plot[Discipline == "Economics", Disc_abbreviation := "Economics"]
dt_plot[Discipline == "Political Science &\nPublic Administration",
        Disc_abbreviation := "Political Science"]
dt_plot[Discipline == "Sociology", Disc_abbreviation := "Sociology"]
dt_plot[Discipline == "Management", Disc_abbreviation := "Management"]
dt_plot[Discipline == "Anthropology &\n Archaeology",
        Disc_abbreviation := "Anthropology"]
dt_plot$Discipline_for_plot <- dt_plot$Discipline
dt_plot[!Discipline %in% core_SSH, Discipline := "Others"]

#Plotting
dt_plot[Discipline %in% core_SSH] %>%
  mutate(label = if_else(Year == max(Year), as.character(Disc_abbreviation),
                        NA_character_)) %>%
  ggplot(mapping = aes(x=Year, y=Prop, col = Discipline)) +
  geom_smooth(se=F,method = 'loess',
             aes(x=Year, y=Prop,col = Discipline, linetype = Discipline),
             show.legend=FALSE) +
  scale_color_manual(name = "Disciplines", breaks = core_SSH, values = colors) +
  scale_linetype_manual(name = "Disciplines", breaks = core_SSH, values = linetypes) +
  # guides(linetype=guide_legend(keywidth = 3, keyheight = 1,nrow=2,byrow=TRUE),
  #        color=guide_legend(keywidth = 3, keyheight = 1,nrow=2,byrow=TRUE)) +
  new_scale_linetype() + # geoms below can use another linetype scale!
  geom_smooth(dt_plot[Discipline == "Others" & Discipline_for_plot != "Arts"],
             mapping = aes(x=Year, y = Prop#, group = Discipline_for_plot
                           , linetype = Discipline_for_plot),
             color = alpha("gray",.4),
             se=F, method = 'loess') +
  scale_linetype_discrete(name="Other Disciplines") +
  theme_minimal() +
  theme(legend.position="bottom", legend.box = "vertical") +
  ylab("CTC (citations toward category)") +
  xlim(c(NA, 2026)) +
  theme(axis.text=element_text(size=13),
        axis.title=element_text(size=14),
        legend.text=element_text(size=13),
        legend.title=element_text(size=14)
        ) +
  guides(linetype=guide_legend(keywidth = 3, keyheight = 1,nrow=4,byrow=TRUE)) +
  geom_label_repel(aes(label = label), nudge_x = 1,
                  xlim = c(2015, NA), na.rm = TRUE, show.legend = FALSE, size= 5) +
  annotate("text",x=max(dt_plot$Year),y=-Inf,hjust=.7,vjust=-0.5,
         label="Curves smoothed by local polynomial regression",
         size=4,color="darkgrey" )

## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'

```



Other Disciplines

- Demography
- Other Social Sciences
- Geography
- Health
- Other Professional Fields
- International R
- Education
- Law
- Humanities

```
ggsave( paste0("figures/", "fig7.png"), width = 16, height = 9, bg = "white",
  scale = .7)
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

```
shades_of_gray <- c("Psychology" = "#404040",
  "Economics" = "black",
  "Sociology" = "#202020",
  "Anthropology & \n Archaeology" = "#A0A0A0",
  "Political Science & \n Public Administration" = "#606060",
  "Management" = "808080",
  "Others" = alpha("gray",.3))
```

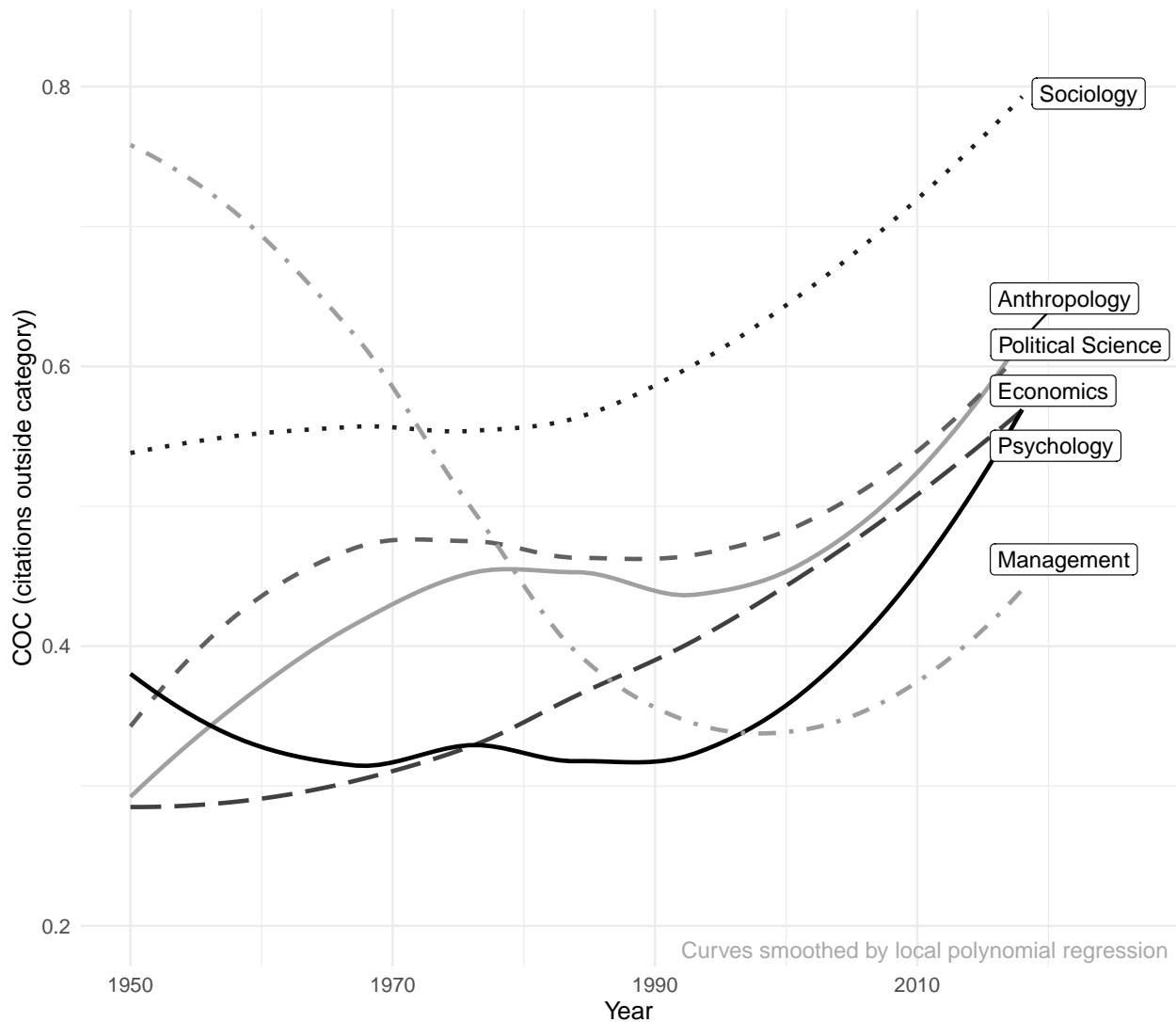
```
dt_plot[Discipline %in% core_SSH] %>%
  mutate(label = if_else(Year == max(Year), as.character(Disc_abbreviation),
    NA_character_)) %>%
  ggplot(mapping = aes(x=Year, y=Prop, gr = Discipline)) +
  geom_smooth(se=F,method = 'loess',
    aes(x=Year, y=Prop,col = Discipline, linetype = Discipline),
    show.legend=FALSE) +
```

```

scale_color_manual(name = "Disciplines", breaks = core_SSH, values = shades_of_gray) +
scale_linetype_manual(name = "Disciplines", breaks = core_SSH, values = linetypes) +
theme_minimal() +
# guides(linetype=guide_legend(keywidth = 3, keyheight = 1,nrow=2,byrow=TRUE),
#        color=guide_legend(keywidth = 3, keyheight = 1,nrow=2,byrow=TRUE)) +
theme(legend.position="bottom", legend.box = "vertical") +
ylab("COC (citations outside category)") +
xlim(c(NA, 2026)) +
# ylim(c(NA, 0.75)) +
theme(axis.text=element_text(size=10),
      axis.title=element_text(size=12)) +
geom_label_repel(aes(label = label), nudge_x = 1, xlim = c(2015, NA),
                na.rm = TRUE, show.legend = FALSE) +
annotate("text",x=max(dt_plot$Year),y=-Inf,hjust=.7,vjust=-0.5,
        label="Curves smoothed by local polynomial regression",
        size=4,color="darkgrey" )

```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
rm(dt_citing_cited,CTC_per_y_and_disc)
```

The same data in a table:

```
dt_plot[order(Discipline,Year),list(Discipline, Year, Prop )]
```

```
##      Discipline Year      Prop
## 1: Sociology 1950 0.5820611
## 2: Sociology 1951 0.4461806
## 3: Sociology 1952 0.5839286
## 4: Sociology 1953 0.4759358
## 5: Sociology 1954 0.5042735
## ---
## 1086:      Others 2018 0.5961835
## 1087:      Others 2018 0.4325403
## 1088:      Others 2018 0.5929338
## 1089:      Others 2018 0.7137917
## 1090:      Others 2018 0.6336377
```

#### 4.4 The balance of toward and outward citations

After having studied separately COC and CTC, we can bring them together in a unique figure:

```
#Disciplines to be plotted
core_SSH <- c("Psychology", "Economics",
              "Political Science &\nPublic Administration",
              "Sociology", "Anthropology &\n Archaeology",
              "Management")

#The table containing every citing-cited pairs
dt_citing_cited <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))
dt_citing_cited <- dt_citing_cited[!is.na(cited_discipline)]

#We merge it with discipline_info which contains discipline names
dt_citing_cited <- merge(dt_citing_cited,
                        discipline_info[,list(Code_Discipline, discipline)],
                        by.x = "cited_discipline", by.y = "Code_Discipline")
setnames(dt_citing_cited, "discipline", "Discipline_cited")
dt_citing_cited <- merge(dt_citing_cited,
                        discipline_info[,list(Code_Discipline, discipline)],
                        by.x = "citing_discipline", by.y = "Code_Discipline")
setnames(dt_citing_cited, "discipline", "Discipline_citing")

fct_balance_ref_graph_w_conf_interval <- function(dt_citing_cited,dt_only=FALSE)
{
  # Creating a new table with needed columns
  dt_p <- dt_citing_cited[!is.na(cited_discipline), list(Year, Discipline_cited,
                                                         Discipline_citing)]

  # Counting how many citations goes from every discipline to every other disciplines
  dt_p <- dt_p[,.N, by = .(Year, Discipline_cited, Discipline_citing)]
  setnames(dt_p, "N", "Nb_cited")

  # Summing total number of citation per discipline citing and per discipline cited
  dt_p[,Total_citing := sum(Nb_cited), by=.(Discipline_citing, Year)]
}
```

```

dt_p[,Total_cited := sum(Nb_cited), by=(Discipline_cited, Year)]

# Table to calculate out proportion
# (what proportion of cited article per a discipline isn't the discipline itself)
prop_out <- dt_p[Discipline_citing != Discipline_cited,
  list(Total_citing = unique(Total_citing), Total_out = sum(Nb_cited)),
  by=(Discipline_citing, Year)]
prop_out$Direction <- "From the discipline (COC)" #For plotting purposes

# Table to calculate toward proportion (what proportion of citation going
# to one discipline doesn't come from the discipline itself)
prop_toward <- dt_p[Discipline_citing != Discipline_cited
  , list(Total_cited = unique(Total_cited),
  Total_toward = sum(Nb_cited)),
  by=(Discipline_cited, Year)]
prop_toward$Direction <- "To the discipline (CTC)"

# Joining both tables and computing balance with confidence interval
setnames(prop_toward,"Discipline_cited", "Discipline")
setnames(prop_out,"Discipline_citing", "Discipline")
balance <- merge(prop_out[,list(Discipline,Year,Total_citing,Total_out)],
  prop_toward[,list(Discipline,Year,Total_cited,Total_toward)],
  by= c("Discipline","Year"))
balance <- balance[Discipline %in% core_SSH] %>% unique()
setkey(balance, Year, Discipline)
balance <- balance[between(Year, 1950, 2018)]
balance[, `:=`(
  low_bound =prop.test(x= c(Total_toward,Total_out), n = c(Total_cited,Total_citing),
    alternative = "two.sided",correct=FALSE)$conf.int[1],
  high_bound =prop.test(x= c(Total_toward,Total_out), n = c(Total_cited,Total_citing),
    alternative = "two.sided",correct=FALSE)$conf.int[2],
  balance = Total_toward/Total_cited - Total_out/Total_citing
  ), by= .(Discipline,Year)]

# Format for ggplot
toward_out <- rbind(
  balance[,list(Discipline,Year, Proportion= Total_toward/Total_cited, balance,
    low_bound, high_bound, Direction= "To the discipline (CTC)"]],
  balance[,list(Discipline,Year, Proportion= -Total_out/Total_citing, balance,
    low_bound, high_bound, Direction= "From the discipline (COC)"]])
)
toward_out$Direction <- factor(toward_out$Direction,
  levels = c("To the discipline (CTC)",
  "From the discipline (COC)"))

#Plotting
if(!dt_only){
  a <- .35 ; color_interval <- "black"
  ggplot(toward_out, aes(x=Year, y=Proportion, fill=Direction)) +
    geom_bar(stat="identity", position="identity") +
    geom_line(aes(x=Year, y=balance,
      color = "Balance (with 95%\nconfidence interval)")) +
    geom_ribbon(aes(ymin=low_bound, ymax=high_bound),fill=color_interval,

```

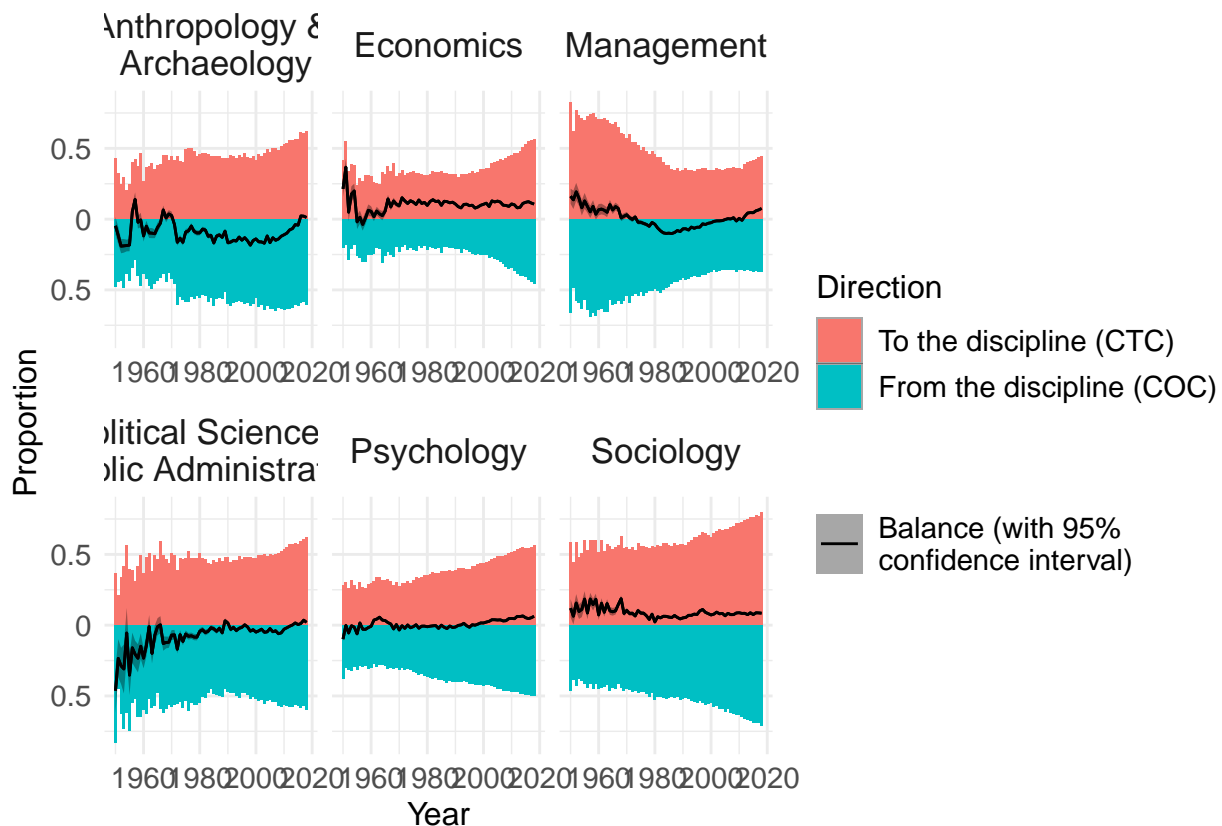


```

    alpha=a) + theme_minimal() +
facet_wrap(~ Discipline, scales = "free_x") +
#ggtitle("Balance of references in the main SSH disciplines") +
scale_color_manual(name = "", values =
  c("Balance (with 95%\nconfidence interval)" = "black")) +
scale_y_continuous(breaks = pretty(toward_out$Proportion),
  labels = abs(pretty(toward_out$Proportion))) +
theme(legend.key = element_rect(colour = "transparent",
  fill = alpha(color_interval, a)),
  axis.text=element_text(size=11),
  axis.title=element_text(size=12),
  legend.text=element_text(size=11),
  legend.title=element_text(size=12),
  strip.text.x = element_text(size = 13))
} else{
  return(toward_out)
}
}

fct_balance_ref_graph_w_conf_interval(dt_citing_cited)

```



```

ggsave( paste0("figures/", "fig8.png"), width = 16, height = 9, bg = "white",
  scale = .6)

rm(dt_citing_cited)

```

And another way to look at this comparison is to take the log ratio:

```

#Disciplines to be plotted
core_SSH <- c("Psychology", "Economics",
             "Political Science &\nPublic Administration",
             "Sociology", "Anthropology &\n Archaeology",
             "Management")

#The table containing every citing-cited pairs
dt_citing_cited <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))
dt_citing_cited <- dt_citing_cited[!is.na(cited_discipline)]

#We merge it with discipline_info which contains discipline names
dt_citing_cited <- merge(dt_citing_cited,
                        discipline_info[,list(Code_Discipline, discipline)],
                        by.x = "cited_discipline", by.y = "Code_Discipline")
setnames(dt_citing_cited, "discipline", "Discipline_cited")
dt_citing_cited <- merge(dt_citing_cited,
                        discipline_info[,list(Code_Discipline, discipline)],
                        by.x = "citing_discipline", by.y = "Code_Discipline")
setnames(dt_citing_cited, "discipline", "Discipline_citing")

toward_out <- fct_balance_ref_graph_w_conf_interval(dt_citing_cited, dt_only = TRUE)

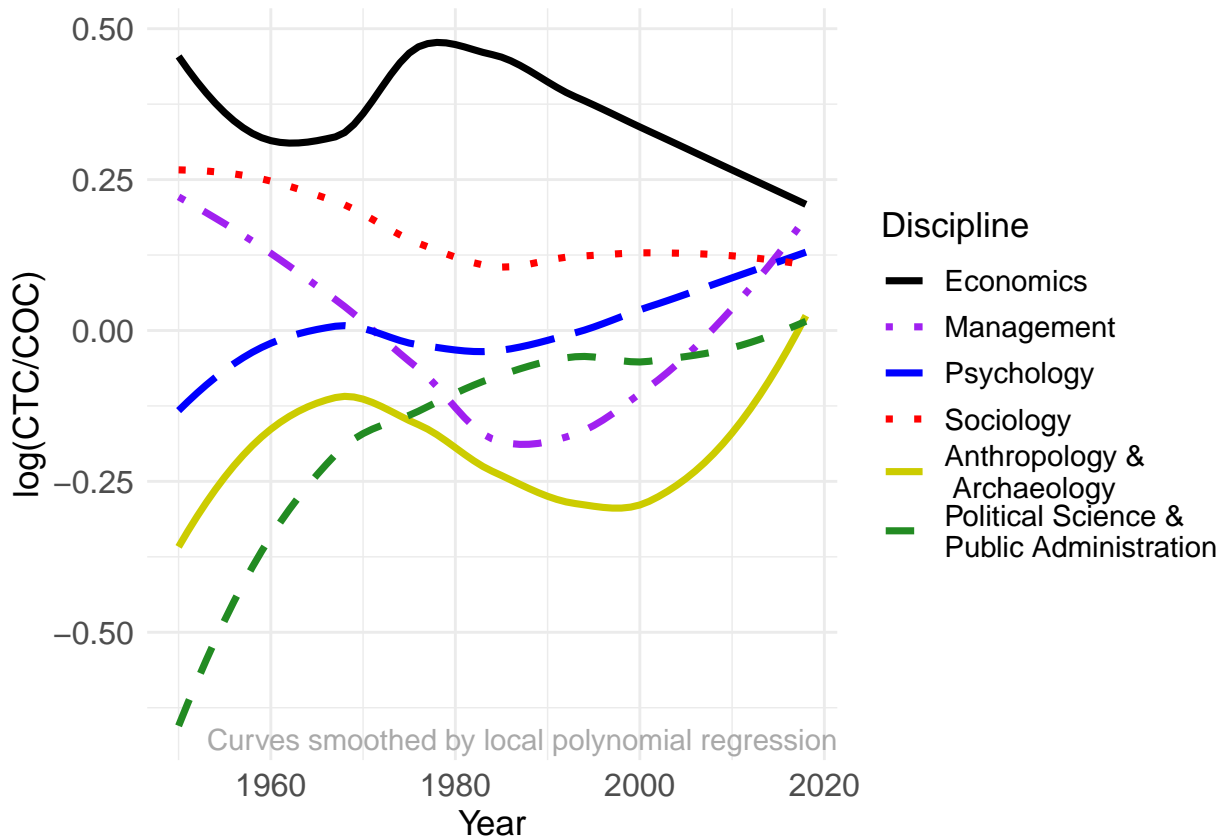
dt_diff <- toward_out[
  ,list(log_diff=log(.SD[Direction=="To the discipline (CTC)", Proportion]/
                    .SD[Direction=="From the discipline (COC)", -Proportion])),
  by=(Year, Discipline)]

# Smoothing y coordinate outside ggplot with loess
dt_diff <- dt_diff[,list(Year, log_diff, log_diff_smooth =
                        predict(loess(log_diff~Year), Year)), by=Discipline]

# Finding the order at the end to order in the legend
order_disc <- dt_diff[Year== max(Year)][order(-log_diff_smooth), Discipline]
dt_diff$Discipline <- factor(dt_diff$Discipline, levels = order_disc)

ggplot(dt_diff, aes(x=Year, y=log_diff_smooth, col=Discipline, linetype=Discipline)) +
  geom_line( lwd=1.25) + scale_color_manual(name = "Discipline", values = colors) +
  theme_minimal() +
  scale_linetype_manual(name = "Discipline", values = linetypes) + ylab("log(CTC/COC)") +
  theme(axis.text=element_text(size=12),
        axis.title=element_text(size=13#, face="bold"
                                ),
        legend.text=element_text(size=11),
        legend.title=element_text(size=13)) +
  annotate( "text", x=max(dt_diff$Year), y=-Inf, hjust=.95, vjust=-0.5,
          label="Curves smoothed by local polynomial regression",
          size=4, color="darkgrey" )

```



```
ggsave( paste0("figures/", "fig9.png"), width = 16, height = 9, bg = "white",
         scale = .6)
```

```
rm(dt_citing_cited, toward_out, dt_diff)
```

## 5 Disaggregating the relations of economics

### 5.1 SSH and NSE

The first graph to make is how the split to investigate is the proportion of citations from the SSH (social sciences and humanities) that goes to the NSE (natural sciences, engineering and biomedical sciences):

```
#Looking at the proportion of cited discipline per year
y_max = max_period
dt_citing_cited <- readRDS(paste0(path_to_project_data,
                                "dt_citing_cited_all.rds"))
dt_citing_cited <- dt_citing_cited[!is.na(cited_discipline) &
                                cited_discipline<200 & # removing discipline unknown
                                between(Year,min_period,max_period)]
discipline_inf <- fread(paste0(path_to_discipline_info, "Liste_Discipline.txt"))
dt_citing_cited <- merge(dt_citing_cited,
                        discipline_inf[,list(Code_Discipline,EGrande_Discipline,
                                             EDiscipline, ESpecialite)],
                        by.x = "cited_discipline", by.y = "Code_Discipline")

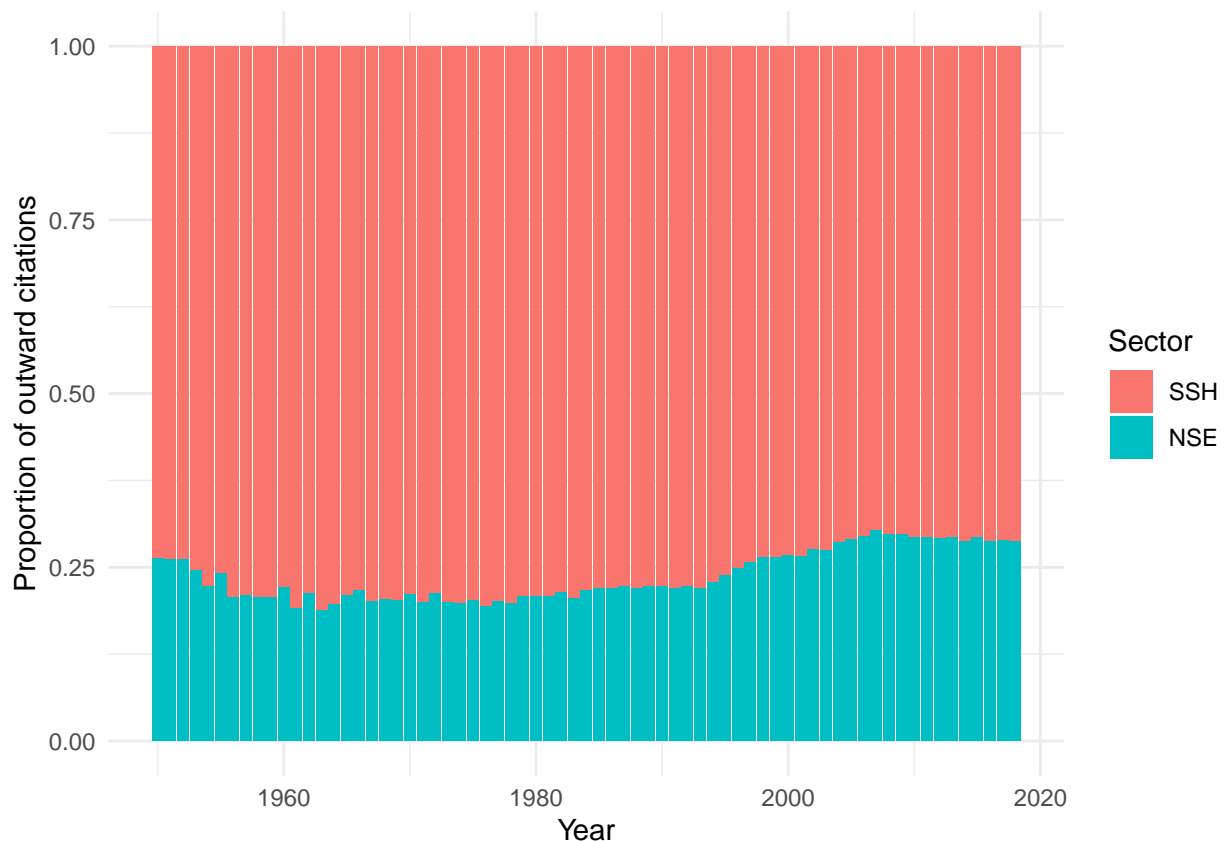
dt_citing_cited[,nb_cited:=.N,by=Year]
```

```

# The proportion of cited articles per EGrande_Discipline
# (Natural Sciences and Engineering and Social Sciences and Humanities)
# among SSH citing articles
prop_grande_disc <- dt_citing_cited[between(citing_discipline,101,200), list(
  EGrande_Discipline = unique(EGrande_Discipline),
  Proportion=.SD[,.(tot=.N),by=EGrande_Discipline]$tot/.N),
  by =Year]

prop_grande_disc <- prop_grande_disc[EGrande_Discipline != "Unknown"]
prop_grande_disc[grep("Natural",EGrande_Discipline),Sector:=factor("NSE")]
prop_grande_disc[grep("Social",EGrande_Discipline),Sector:=as.factor("SSH")]
#order_disc <- unique(prop_grande_disc[Year == y_max,][order(Proportion),Sector])
prop_grande_disc$Sector <- factor(prop_grande_disc$Sector, levels = c( "SSH", "NSE"))
ggplot(prop_grande_disc, aes(x = Year, y = Proportion, fill = Sector)) +
  geom_bar(stat = "identity") +
  labs(y="Proportion of outward citations") + theme_minimal() #+

```



```

# theme(axis.text=element_text(size=12),
#       axis.title=element_text(size=13#,face="bold"
#                                ),
#       legend.text=element_text(size=11),
#       legend.title=element_text(size=13))
#+ ggtitle("Proportion of extradisciplinary citations : SSH vs NSE")

ggsave( paste0("figures/", "fig10a.png"), width = 8, height = 6, bg = "white",
  scale = .6 )

```

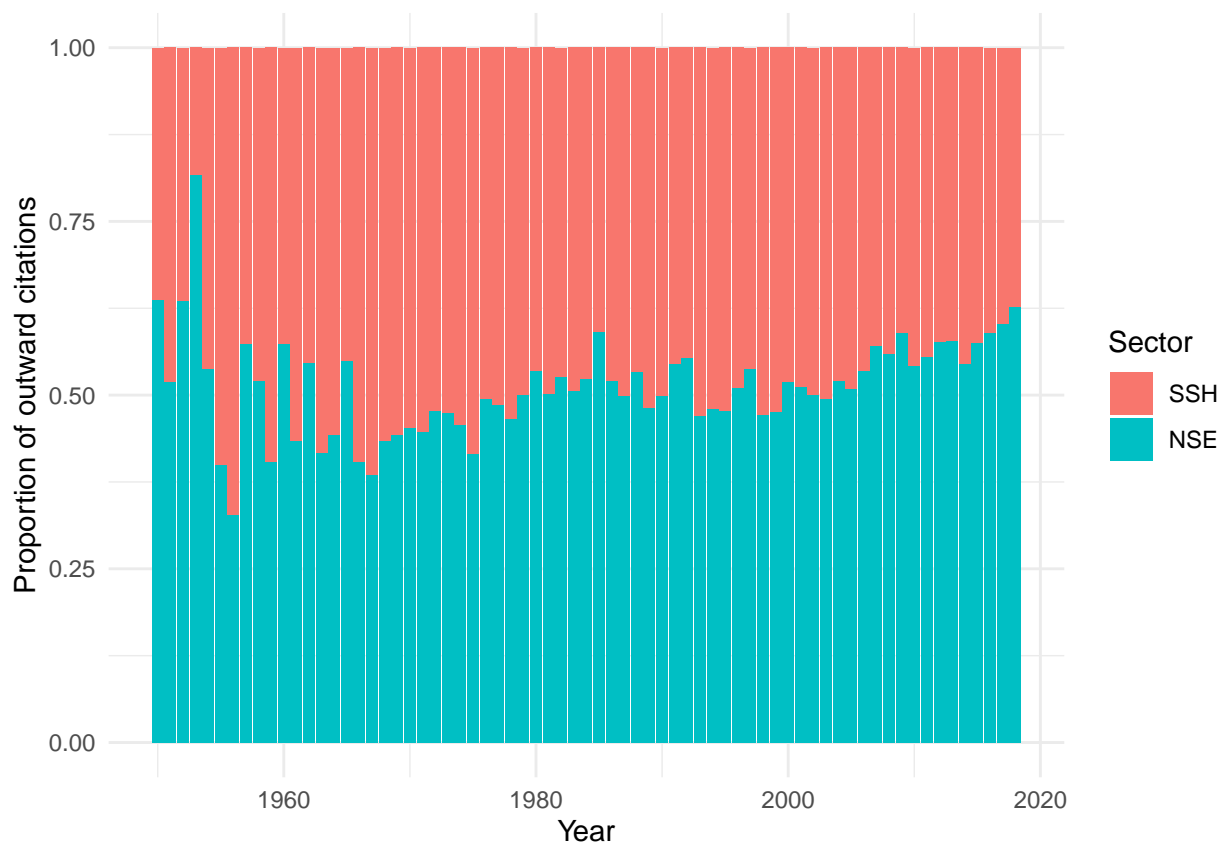
```

##Doing the same thing with economics
dt_citing_cited <- dt_citing_cited[citing_discipline == 119 &
  cited_discipline != 119 &
  cited_discipline != 132,]

##SSH vs NSE in economics
prop_grande_disc_econ <- dt_citing_cited[, nb_cited := .N, by = "Year"] [
  , nb_cited_g_disc := .N, by = .(Year, EGrande_Discipline)] [
  , Proportion := nb_cited_g_disc / nb_cited] [
  , list(Year, EGrande_Discipline, Proportion)] %>% unique()
prop_grande_disc_econ <- prop_grande_disc_econ[EGrande_Discipline != "Unknown"]
prop_grande_disc_econ[grepl("Natural", EGrande_Discipline), Sector:="NSE"]
prop_grande_disc_econ[grepl("Social", EGrande_Discipline), Sector:="SSH"]
order_disc <- unique(prop_grande_disc_econ[Year == y_max,][order(Proportion),Sector])
prop_grande_disc_econ$Sector <- factor(prop_grande_disc_econ$Sector,
  levels = c("SSH", "NSE"))

ggplot(prop_grande_disc_econ, aes(x = Year, y = Proportion, fill = Sector)) +
  geom_bar(stat="identity") +
  labs(y="Proportion of outward citations") + theme_minimal()

```



```

#+ ggtitle("Proportion of extradisciplinary citations for economics : SSH vs NSE")

ggsave( paste0("figures/", "fig10b.png"), width = 8, height = 6, bg = "white",
  scale = .6 )

```

```
rm(dt_citing_cited)
```

Through time, the average proportion going to NSE is:

- 0.24 for all SSH
- 0.51 for all SSH

For all the SSH, the proportion going to NSE stopped growing in 2007.

## 5.2 Strong ties of economics

Now, focussing on economics, we can find the disciplines to which the highest shares of its citations have gone over the studied period:

```
the_disc <- "Economics"
the_varia_SSH <- "Others SSH"
the_varia_NSE <- "Others NSE"
SSH <- "Social Sciences and Humanities"
NSE <- "Natural Sciences and Engineering"

#Data needed
dt_citing_cited <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))
dt_citing_cited <- dt_citing_cited[between(Year, 1950, 2019),]

#Summing how many citation goes from every discipline to every other discipline per year
citing_cited_disc <- dt_citing_cited[, list(Citations =.N),
                                         by=c("Year", "citing_discipline",
                                                "cited_discipline")]

#Merging with discipline_info to get cited discipline names
citing_cited_disc <- merge(citing_cited_disc,
                          discipline_info[,list(EGrande_Discipline,
                                                  Code_Discipline,
                                                  Discipline_cited = discipline)],
                          by.x="cited_discipline", by.y="Code_Discipline", all.x=T)

#Merging with discipline_info to get citing discipline names
citing_cited_disc <- merge(citing_cited_disc,
                          discipline_info[,list(Code_Discipline,
                                                  Discipline_citing = discipline)],
                          by.x="citing_discipline", by.y="Code_Discipline", all.x=T)

#Getting references for Economics only
cit_disc <- citing_cited_disc[Discipline_citing == the_disc & !is.na(cited_discipline)]

#Summing citations per discipline cited
cit_disc <- cit_disc[,list(Citations= sum(Citations)),
                     by=list(Year,Discipline_citing,
                              Discipline_cited, EGrande_Discipline)]

setkey(cit_disc,Year)

#Calculating proportion and taking relevant columns
cit_disc <- cit_disc[, list(EGrande_Discipline, Discipline_cited,
                           Proportion = Citations/sum(Citations)),by=Year]
```

```

# Finding out the most important disciplines:
imp_disc <- cit_disc[Discipline_cited != the_disc, list(highest_prop = max(Proportion)),
  by=Discipline_cited][order(-highest_prop)][
  1:n_disc, Discipline_cited]
dt_p <- cit_disc[Discipline_cited %in% imp_disc, list(Year, Discipline_cited, Proportion)]

#Adding a all_others column for plotting purposes
dt_p$all_others <- c("Cited Disciplines")

#Agregating all other cited discipline and dividing it in Other SSH and Other NSE
dt_p_all_others <- cit_disc[!Discipline_cited %in% c(imp_disc,the_disc) &
  EGrande_Discipline == SSH,
  list(Discipline_cited=the_varia_SSH,
  Proportion= sum(Proportion)),by=Year]
dt_p_all_others <- rbind(dt_p_all_others,
  cit_disc[!Discipline_cited %in% c(imp_disc,the_disc) &
  EGrande_Discipline == NSE,
  list(Discipline_cited=the_varia_NSE,
  Proportion= sum(Proportion)),by=Year])

dt_p_all_others[, all_others := Discipline_cited]

#Merging important discipline with all others
dt_p <- (rbind(dt_p, dt_p_all_others))

#Colors for every disciplines.
colors_others <- c("Earth and Space" = alpha("brown", 0.7),
  "Sociology" = "#FF0000",
  "Mathematics" = alpha("darkgreen", 0.6),
  "Biomedical Research" = alpha("#00e5e5", 0.7),
  "Management" = "purple",
  "Biology" = alpha("green", 0.7),
  "Other Engineering \n and Technology" = alpha("#E59400", 0.7))

#Linetypes for every disciplines.
linetypes_others <- c("Earth and Space" = "solid",
  "Sociology" = "dotted",
  "Mathematics" = "solid",
  "Biomedical Research" = "solid",
  "Management" = "dotdash",
  "Biology" = "solid",
  "Other Engineering \n and Technology" = "solid")

#Ordering for plotting purposes
order_disc <- unique(dt_p[, Prop := predict(loess(Proportion~Year), Year),
  by = Discipline_cited][Year == 2018,][
  order(-Prop),Discipline_cited])
dt_p$Discipline_cited <- factor(dt_p$Discipline_cited, levels = order_disc)

ggplot(dt_p[all_others == "Cited Disciplines"]) +
  geom_smooth(se=F,method = 'loess',aes(x=Year, y=Proportion,col = Discipline_cited,

```

```

                                linetype = Discipline_cited)) +
scale_color_manual(name = "Disciplines", values = colors_others) +
scale_linetype_manual(name = "Disciplines", values = linetypes_others) +
guides(linetype=guide_legend(keywidth = 1.5, keyheight = 1,byrow=TRUE),
       color=guide_legend(keywidth = 1.5, keyheight = 1,byrow=TRUE)) +
new_scale_linetype() + # geoms below can use another linetype scale!
geom_smooth(dt_p[all_others != "Cited Disciplines"],
            mapping = aes(x=Year, y = Proportion,#, group = Discipline_for_plot
                          linetype = all_others),
            color = "gray",
            se=F, method = 'loess') +
scale_linetype_manual(name="Other Disciplines",
                      values = c("Others NSE" = "longdash",
                                  "Others SSH" = "dashed")) +
#theme(legend.box = "horizontal") +
ylab("COC (citations outside category)") + theme_minimal() +
theme(legend.position="bottom", legend.box = "vertical") +
theme(axis.text=element_text(size=13),
      axis.title=element_text(size=14),
      legend.text=element_text(size=13),
      legend.title=element_text(size=14)
      ) +
annotate( "text",x=max(dt_p$Year),y=-Inf,hjust=.95,vjust=-0.5,
          label="Curves smoothed by local polynomial regression",
          size=4,color="darkgrey" )

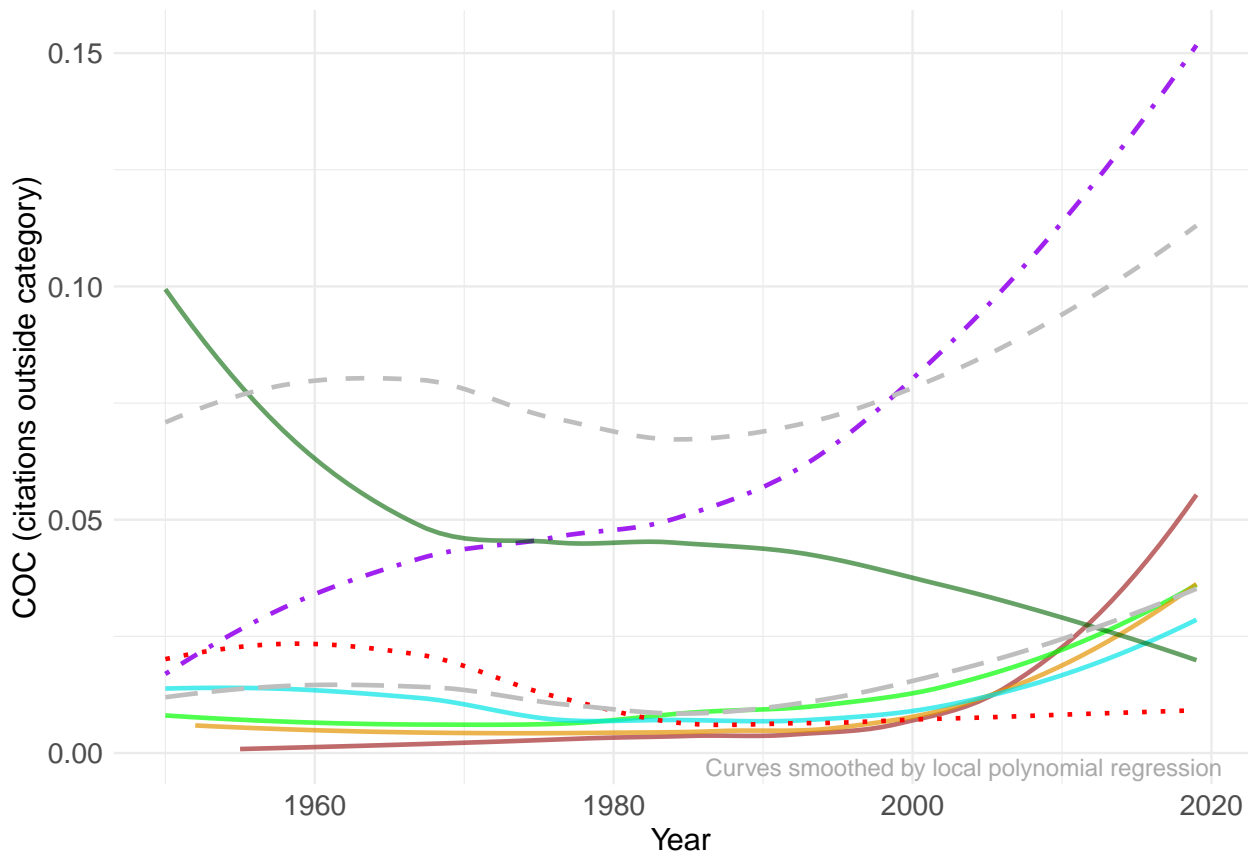
```

```

## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'

```





Disciplines

- Management
- Earth and Space
- Biology
- Other Engine and Technol
- Biomedical Research
- Mathematics
- Sociology
- Other Disciplines
- Others NSE
- Others SSH

```
ggsave( paste0("figures/", "fig11.png"), width = 16, height = 9, bg = "white",
        scale = .7)
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
rm(dt_citing_cited); rm(citing_cited_disc)
```

Information about mathematics:

- Highest share of 0.15 occurred in year 1952.
- Lowest point is at 0.02 in year 2019.

Information about management:

- Highest share of 0.145 occurred in year 2014.
- Lowest point is at 0.005 in year 1950.
- In the last five years of our period (2014 to 2018), the average proportion of the extradisciplinary citations that go to management is 0.331 (i.e., 0.141 to 0.428)

And here are the data in a table

```
dt_p[all_others == "Cited Disciplines"][order(Discipline_cited, Year),
list(Discipline_cited, Year,Proportion)]
```

```
##      Discipline_cited Year  Proportion
##  1:      Management 1950 0.004987531
##  2:      Management 1951 0.012861736
##  3:      Management 1952 0.026385224
##  4:      Management 1953 0.022522523
##  5:      Management 1954 0.024340771
## ---
## 464:      Sociology 2015 0.008240946
## 465:      Sociology 2016 0.008912295
## 466:      Sociology 2017 0.008938363
## 467:      Sociology 2018 0.009322987
## 468:      Sociology 2019 0.007821973
```

### 5.2.1 Weighed citations

It is possible to produce the same figure but weighing citations based on the journal where they originate.

```
the_disc <- "Economics"
the_varia_SSH <- "Others SSH"
the_varia_NSE <- "Others NSE"
SSH <- "Social Sciences and Humanities"
NSE <- "Natural Sciences and Engineering"

dt_citing_cited_journals <- readRDS(paste0(path_to_project_data,
"dt_citing_cited_journals.rds"))

load(paste0(path_to_project_data, "journal_impact_per_window.RData"))
setnames(dt_citing_cited_journals, "Year", "citing_Year")

# Focus on econ
cit_disc <- dt_citing_cited_journals[Discipline_citing == the_disc &
!is.na(Discipline_cited),
list(citing_code_revue, Discipline_citing,
Discipline_cited,
citing_Year, N)][
, list(Nb_cited =sum(N)),
by = list(citing_code_revue,
Discipline_cited,
citing_Year) ]

# Getting the weights in there.
cit_disc <- merge(cit_disc,
journal_impact_per_window[
, list(journal_ID, citing_Year,
Journal_weight_in_disc =impact_weight )],
by.x = c("citing_code_revue", "citing_Year"),
by.y = c("journal_ID", "citing_Year"))

## Calculating proportions of extradisciplinary citations
```

```

#Proportion of extradisciplinary citations to various disciplines:
setkey(cit_disc,citing_Year)
cit_disc[,tot_weighted_cit:= sum(Nb_cited*Journal_weight_in_disc),by=citing_Year]
cit_disc <- cit_disc[, list( Proportion = sum(Nb_cited*Journal_weight_in_disc)/
                           unique(tot_weighted_cit)
                           ),
                           by =.(Discipline_cited, citing_Year)]
setnames(cit_disc,"citing_Year","Year")

# Merging with discipline info to get EGrande_Disc
cit_disc <- merge(cit_disc,
                 unique(discipline_info[,list(EGrande_Discipline,
                                               Discipline_cited = discipline)]),
                 by= "Discipline_cited", all.x=T)

# Finding out the most important disciplines:
imp_disc <- cit_disc[Discipline_cited != the_disc, list(highest_prop = max(Proportion)),
                   by=Discipline_cited][order(-highest_prop)][
                   1:n_disc, Discipline_cited]
dt_p <- cit_disc[Discipline_cited %in% imp_disc, list(Year, Discipline_cited, Proportion)]

#Adding a all_others column for plotting purposes
dt_p$all_others <- c("Cited Disciplines")

#Agregating all other cited discipline and dividing it in Other SSH and Other NSE
dt_p_all_others <- cit_disc[!Discipline_cited %in% c(imp_disc,the_disc) &
                          EGrande_Discipline == SSH,
                          list(Discipline_cited=the_varia_SSH,
                                Proportion= sum(Proportion)),by=Year]
dt_p_all_others <- rbind(dt_p_all_others,
                        cit_disc[!Discipline_cited %in% c(imp_disc,the_disc) &
                                EGrande_Discipline == NSE,
                                list(Discipline_cited=the_varia_NSE,
                                      Proportion= sum(Proportion)),by=Year])

dt_p_all_others[, all_others := Discipline_cited]

#Merging important discipline with all others
dt_p <-(rbind(dt_p, dt_p_all_others))

#Colors for every disciplines.
colors_others <- c("Biomedical Research" = alpha("#00e5e5", 0.7),
                  "Earth and Space" = alpha("brown", 0.7),
                  "Geography" = alpha("#BABA45", 0.5),
                  # "Law" = alpha("#C8A2C8", 0.5),
                  "Other Social Sciences" = alpha("#E59400", 0.5),
                  "Management" = "purple",
                  "Mathematics" = alpha("darkgreen", 0.6),
                  "Other Engineering \n and Technology" = alpha("#E59400", 0.7)#,
                  #
                  "Psychology" = "#0000FF"
                  )

```

```

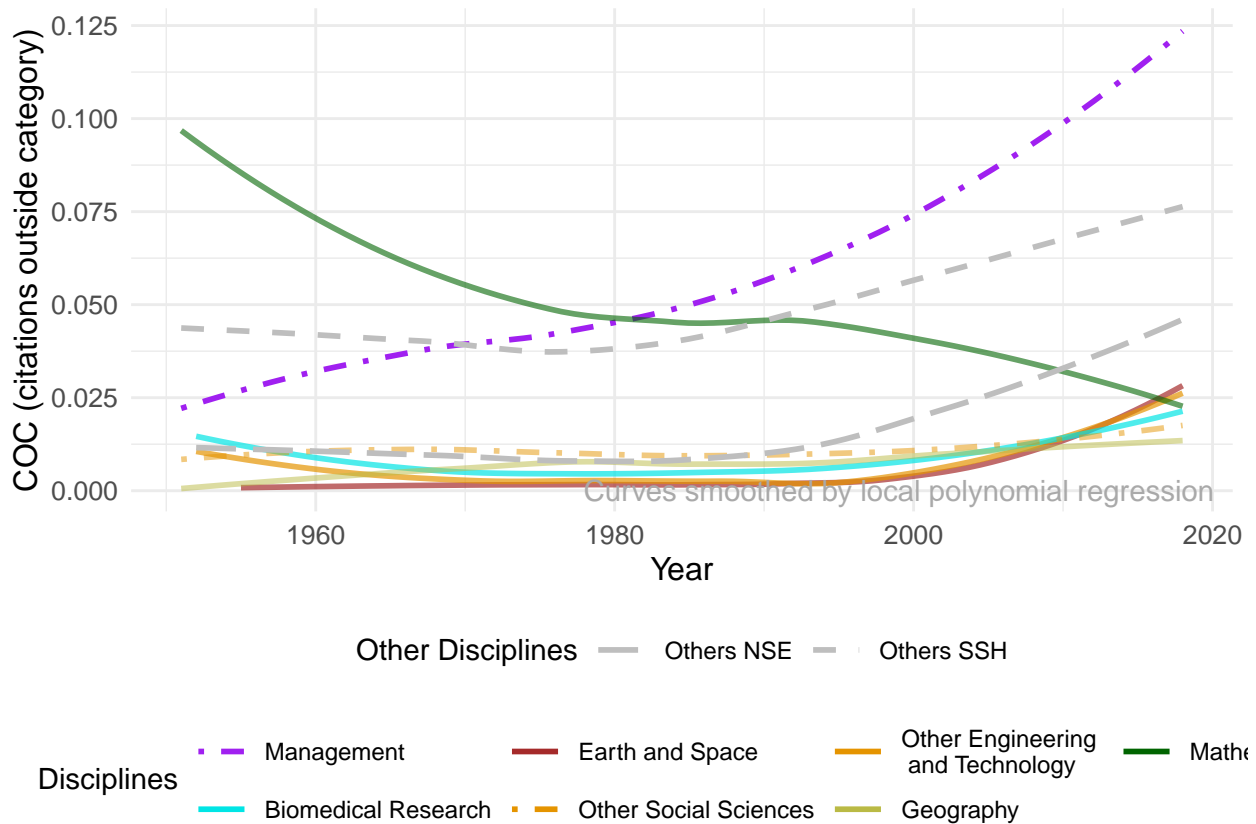
#Linetypes for every disciplines.
linetypes_others <- c("Biomedical Research" = "solid",
  "Earth and Space" = "solid",
  "Geography" = "solid",
  # "Law" = "longdash",
  "Other Social Sciences" = "dotdash",
  "Management" = "dotdash",
  "Mathematics" = "solid",
  "Other Engineering \n and Technology" = "solid"#,
  # "Psychology" = "longdash"
)

#Ordering for plotting purposes
order_disc <- unique(dt_p[, Prop := predict(loess(Proportion~Year), Year),
  by = Discipline_cited][Year == 2018,][
  order(-Prop),Discipline_cited])
dt_p$Discipline_cited <- factor(dt_p$Discipline_cited, levels = order_disc)

ggplot(dt_p[all_others == "Cited Disciplines"]) +
  geom_smooth(se=F,method = 'loess',
    aes(x=Year, y=Proportion,col = Discipline_cited,
      linetype = Discipline_cited)) +
  scale_color_manual(name = "Disciplines", values = colors_others) +
  scale_linetype_manual(name = "Disciplines", values = linetypes_others) +
  guides(linetype=guide_legend(keywidth = 1.5, keyheight = 1,byrow=TRUE),
    color=guide_legend(keywidth = 1.5, keyheight = 1,byrow=TRUE)) +
  new_scale_linetype() + # geoms below can use another linetype scale!
  geom_smooth(dt_p[all_others != "Cited Disciplines"],
    mapping = aes(x=Year, y = Proportion, #, group = Discipline_for_plot
      linetype = all_others),
    color = "gray",
    se=F, method = 'loess') +
  scale_linetype_manual(name="Other Disciplines",
    values = c("Others NSE" = "longdash",
      "Others SSH" = "dashed")) +
  #theme(legend.box = "horizontal") +
  ylab("COC (citations outside category)") + theme_minimal() +
  theme(legend.position="bottom", legend.box = "vertical") +
  theme(axis.text=element_text(size=10),
    axis.title=element_text(size=12#,face="bold"
    )) +
  annotate( "text",x=max(dt_p$Year),y=-Inf,hjust=.95,vjust=-0.5,
    label="Curves smoothed by local polynomial regression",
    size=4,color="darkgrey" )

## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'

```



```
rm(dt_citing_cited); rm(citing_cited_disc)
```

```
## Warning in rm(dt_citing_cited): object 'dt_citing_cited' not found
```

```
## Warning in rm(citing_cited_disc): object 'citing_cited_disc' not found
```

We see that the patterns are similar with weighed and unweighed citations.

### 5.2.2 A detour to understand management better

Since management has become so important for economics, we want to understand a bit more its own bibliometric history. We disaggregate its strong ties:

```
the_disc <- "Management"
the_varia_SSH <- "Others SSH"
the_varia_NSE <- "Others NSE"
SSH <- "Social Sciences and Humanities"
NSE <- "Natural Sciences and Engineering"

#Data needed
dt_citing_cited <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))
dt_citing_cited <- dt_citing_cited[between(Year, 1950, 2019),]

#Summing how many citation goes from every discipline to every other discipline per year
citing_cited_disc <- dt_citing_cited[, list(Citations =.N),
                                     by=c("Year", "citing_discipline",
                                           "cited_discipline")]

#Merging with discipline_info to get cited discipline names
citing_cited_disc <- merge(citing_cited_disc,
```

```

        discipline_info[,list(EGrande_Discipline,
                              Code_Discipline,
                              Discipline_cited = discipline)],
        by.x="cited_discipline", by.y="Code_Discipline", all.x=T)
#Merging with dicipline_info to get citing discipline names
citing_cited_disc <- merge(citing_cited_disc,
                          discipline_info[,list(Code_Discipline,
                                                  Discipline_citing = discipline)],
                          by.x="citing_discipline", by.y="Code_Discipline", all.x=T)

#Getting references for Economics only
cit_disc <- citing_cited_disc[Discipline_citing == the_disc & !is.na(cited_discipline)]

#Summing citations per dicipline cited
cit_disc <- cit_disc[,list(Citations= sum(Citations)),
                      by=list(Year,Discipline_citing,Discipline_cited,
                              EGrande_Discipline)]

setkey(cit_disc,Year)

#Calculating proportion and taking relevant columns
cit_disc <- cit_disc[, list(EGrande_Discipline, Discipline_cited,
                           Proportion = Citations/sum(Citations)),by=Year]

# Finding out the most important disciplines:
imp_disc <- cit_disc[Discipline_cited != the_disc, list(highest_prop = max(Proportion)),
                    by=Discipline_cited] [
  order(-highest_prop)][1:n_disc, Discipline_cited]
dt_p <- cit_disc[Discipline_cited %in% imp_disc, list(Year, Discipline_cited,
                                                    Proportion)]

#Adding a all_others column for plotting purposes
dt_p$all_others <- c("Cited Disciplines")

#Agregating all other cited discipline and dividing it in Other SSH and Other NSE
dt_p_all_others <- cit_disc[!Discipline_cited %in% c(imp_disc,the_disc) &
                           EGrande_Discipline == SSH,
                           list(Discipline_cited = the_varia_SSH,
                                 Proportion = sum(Proportion)), by = Year]
dt_p_all_others <- rbind(dt_p_all_others,
                        cit_disc[!Discipline_cited %in% c(imp_disc,the_disc) &
                                 EGrande_Discipline == NSE,
                                 list(Discipline_cited = the_varia_NSE,
                                       Proportion = sum(Proportion)),by = Year])

dt_p_all_others[, all_others := Discipline_cited]

#Merging important discipline with all others
dt_p <- (rbind(dt_p, dt_p_all_others))

#Colors for every disciplines.
colors_others <- c("Mathematics" = alpha("darkgreen", 0.6),
                  "Psychology" = "#0000FF",

```

```

    "Other Social Sciences" = alpha("#E59400", 0.5),
    "Economics" = "black",
    "Sociology" = "#FF0000",
    "Computers & \n Operations Research" = alpha("#00e5e5", 0.7),
    "Education" = "#e75480"
  )

# Linetypes for every disciplines.
# Linetypes for every disciplines. We used Angrist's linetype
# and added dotdash for Management
linetypes_others <- c("Mathematics" = "solid",
  "Psychology" = "longdash",
  "Other Social Sciences" = "dotdash",
  "Economics" = "solid",
  "Sociology" = "dotted",
  "Computers & \n Operations Research" = "dashed",
  "Education" = "dotdash"
)

#Ordering for plotting purposes
order_disc <- unique(dt_p[, Prop := predict(loess(Proportion~Year), Year),
  by = Discipline_cited][Year == 2018,][
  order(-Prop),Discipline_cited])
dt_p$Discipline_cited <- factor(dt_p$Discipline_cited, levels = order_disc)

ggplot(dt_p[all_others == "Cited Disciplines"]) +
  geom_smooth(se=F,method = 'loess',
    aes(x=Year, y=Proportion,col = Discipline_cited,
      linetype = Discipline_cited)) +
  scale_color_manual(name = "Disciplines", values = colors_others) +
  scale_linetype_manual(name = "Disciplines", values = linetypes_others) +
  guides(linetype=guide_legend(keywidth = 1.5, keyheight = 1,byrow=TRUE),
    color=guide_legend(keywidth = 1.5, keyheight = 1,byrow=TRUE)) +
  new_scale_linetype() + # geoms below can use another linetype scale!
  geom_smooth(dt_p[all_others != "Cited Disciplines"],
    mapping = aes(x=Year, y = Proportion,#, group = Discipline_for_plot
      linetype = all_others),
    color = "gray",
    se=F, method = 'loess') +
  scale_linetype_manual(name="Other Disciplines",
    values = c("Others NSE" = "longdash",
      "Others SSH" = "dashed")) +
  #theme(legend.box = "horizontal") +
  ylab("COC (citations outside category)") + theme_minimal() +
  theme(legend.position="bottom", legend.box = "vertical") +
  theme(axis.text=element_text(size=13),
    axis.title=element_text(size=14),
    legend.text=element_text(size=13),
    legend.title=element_text(size=14)
  )

```

```

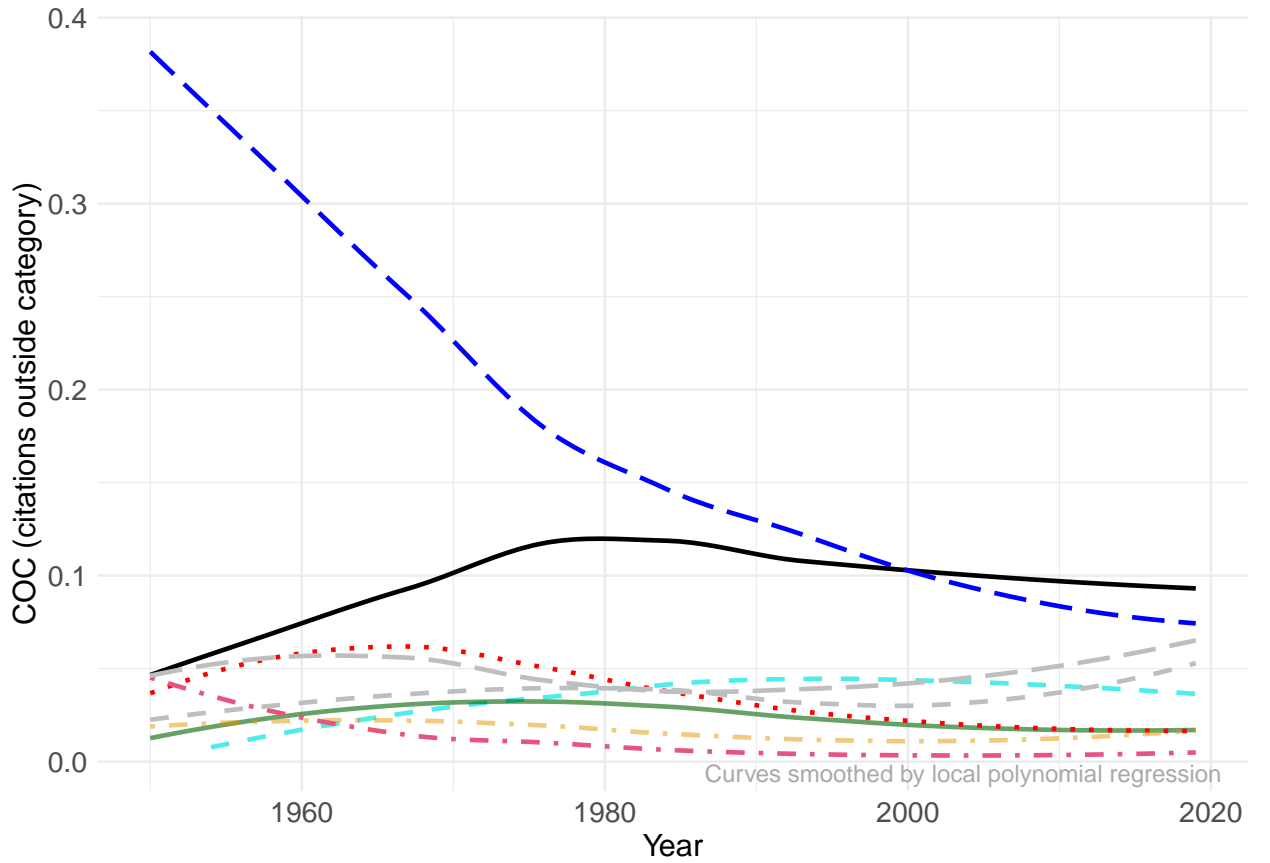
) +
  annotate( "text",x=max(dt_p$Year),y=-Inf,hjust=.95,vjust=-0.5,
           label="Curves smoothed by local polynomial regression",
           size=4,color="darkgrey" )

```

```

## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'

```



Disciplines

- Economics
- Psychology
- Computers & Operations Research
- Mathen
- ... Sociology
- ... Other Social Sciences
- ... Education
- Other Disciplines — Others NSE
- Others SSH

```

ggsave( paste0("figures/", "fig12.png"), width = 16, height = 9, bg = "white",
         scale = .7)

```

```

## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
rm(dt_citing_cited); rm(citing_cited_disc)

```

The link to psychology early in the period is especially strong:

- In the 1950s, the proportion of citations from management that go to psychology averaged at 0.354.

And the table with the data:



```
dt_p[all_others == "Cited Disciplines"][
  order(Discipline_cited, Year), list(Discipline_cited, Year, Proportion)]
```

```
##      Discipline_cited Year  Proportion
##  1:      Economics 1950 0.067500000
##  2:      Economics 1951 0.051948052
##  3:      Economics 1952 0.082066869
##  4:      Economics 1953 0.066276803
##  5:      Economics 1954 0.083657588
## ---
## 482:      Education 2015 0.004090413
## 483:      Education 2016 0.004226854
## 484:      Education 2017 0.004430982
## 485:      Education 2018 0.004377436
## 486:      Education 2019 0.004711638
```

One issue with the general picture drawn so far of economics becoming more outward looking and significantly tightening its relationship with management is that the latter is a pretty diverse field. Economics might be getting closer to some part of Management – e.g., finance – but not to others – e.g., human resources.

We thus split journals in Management between the finance journal and “other management”. We rely on two sources to classify journals:

1. The journal categories of Scimago. More specifically, we extracted the 2018 list of journals in the category ‘Business, Management and Accounting’.
2. The 2019 “Categorization of Journals in Economics and Management” from the French CNRS.

We carefully match the journals in this list to our own journal list and it gives us the following set of finance journals:

```
# Small function to clean up title of journals to maximize matching
clean_text <- function(textvector){
  textvector <- toupper(textvector)
  textvector <- gsub(pattern = "&",
                    replacement = "AND",textvector) #changing all & for "AND"
  textvector <- gsub(pattern = ",",replacement = "",textvector) # removing all commas
  textvector <- gsub(pattern = "- ",replacement = " ",textvector) # removing -
  textvector <- gsub(pattern = "- ",replacement = " ",textvector) # removing -
  textvector <- gsub(pattern = ":",replacement = " ",textvector) # removing :
  textvector <- gsub(pattern = ":",replacement = " ",textvector) # removing :
  textvector <- gsub(pattern = "Û",replacement = "U",textvector) # removing Û

  return(textvector)
}

# Loading Scimago journal list
list_j_scimago <- read.csv(
  paste0(path_to_project_data,
         "2020-03-09_scimagojr_2018_Business_Management_Accounting.csv"),
  sep = ";") %>% data.table()
list_j_scimago <- list_j_scimago[,list(
  Title = clean_text(textvector=Title),
  Categories = as.character(Categories))] # removing many irrelevant columns.
setkey(list_j_scimago,Title)
# There are duplicates.
```

```

# We keep one row per title which is selected based on the line with most categories.
list_j_scimago$nb_catego <- lapply(list_j_scimago$Categories, function(x)
  strsplit(x,split = ";") %>% unlist() %>% length() %>% unlist()
list_j_scimago <- list_j_scimago[,.SD[which.max(nb_catego)][1],by=Title]
list_j_scimago$nb_catego <- NULL

# saving the number of finance journals in this list:
nb_fin_j <- data.table(source= "Scimago",
  nb_fin_journals =
    list_j_scimago[grepl("Finance \\(Q",Categories)] %>%
    nrow() )

# Loading CNRS journal list for Finance and Insurance
list_j_CNRS_finance <- read.csv(
  paste0(path_to_project_data,
    "2020-03-10_CNRS_journals_finance_and_insurance.csv"),
    sep=";") %>% data.table()

list_j_CNRS_finance[,Title := trimws(Title) %>% toupper() %>% clean_text()]
nb_fin_j <- rbind(nb_fin_j,
  data.table(source= "CNRS",
    nb_fin_journals = nrow(list_j_CNRS_finance) ))

# Fetching our list of journals
ESH <- dbConnect(MySQL(), user=usr, password=pswd, dbname='OST_Expanded_SciHum',
  host='127.0.0.1')
disc_w_journal <- dbGetQuery(ESH, "SELECT Code_Revue as journal_ID, Code_Discipline,
  Revue as Title
  FROM OST_Expanded_SciHum.Revues;") %>%
  data.table
disc_w_journal$Title <- strsplit(disc_w_journal$Title ,"\r") %>% unlist() %>%
  clean_text()

# Restricting the focus to the journals in management and econ
disc_w_journal <- disc_w_journal[Code_Discipline %in% c(119,132)]

# Matching procedures ####

# For the CNRS list

# step A: perfect match
disc_w_journal[Title %in% list_j_CNRS_finance$Title, Finance:=TRUE]

# This line (commented out) checks the ranks of the journals in the CNRS list
# that don't match a journal in WoS.
#list_j_CNRS_finance[!Title %in% disc_w_journal[Finance == TRUE,Title]]
# We see that all rank #1 are identified and
# only one journal in rank #2 is not found
# (Journal "Finance" of the French Finance Association, which is indeed not on WoS)
# It thus looks good, but it is perhaps a bit too restrictive
# because the journal list is short.
# We thus also try to extract information from the Scimago list

```

```

# step B: perfect match (for Scimago)
disc_w_journal <- merge(disc_w_journal, list_j_scimago,by="Title",
                        all.x = TRUE,all.y = FALSE)
#removing those from list of journals:
#list_j_scimago <- list_j_scimago[!Title %in% disc_w_journal$Title]

# step C: searching for 'financ', "bank", "credit" in remaining journal names
disc_w_journal[is.na(Categories) & is.na(Finance) & grepl("FINAN",Title),
               Categories:="Finance"]
#disc_w_journal[is.na(Categories) & is.na(Finance) & grepl("BANK",Title),
#               Categories:="Finance"]
disc_w_journal[is.na(Categories) & is.na(Finance) & grepl("RISK",Title),
               Categories:="Finance"]
# manually going through the remaining journal names, it seems to be fine now.

# step D: partial match using fuzzyjoin
# (NOT USED IN THE END because there are false matches and the steps above work well)
# if(!require(fuzzyjoin)){ install.packages("fuzzyjoin"); library(fuzzyjoin) }
# J_partial_match <- stringdist_inner_join(disc_w_journal[is.na(Categories) &
#               Code_Discipline == 132],
#               list_j_scimago, by="Title")
# J_partial_match[,list(Title.x,Title.y)]
# disc_w_journal <- merge(disc_w_journal,J_partial_match[,
#               list(Title =Title.x, Categories.y)], by="Title",
#               all.x = TRUE,all.y = FALSE)
# disc_w_journal[is.na(Categories), Categories := Categories.y][,Categories.y:= NULL]

## step E: Adding to the finance category journals that are detected as finance
# in (more permissive) steps B and C,
# but manually adding constraints such as not also being
# tagged as Accounting, Sports Science, etc.
# This procedure gives an extra 33 journals,
# which are manually checked for their relevance to the field of finance.

disc_w_journal[is.na(Finance) &
               (grepl("Finance \\(Q",Categories) | Categories == "Finance") &
                (grepl("Accounting \\(Q1",Categories)) &
                 (!grepl("Sports Science \\(Q",Categories)) &
                  (!grepl("Management \\(Q",Categories)) &
                   (!grepl("PUBLIC FINANCE",Title)), Finance := TRUE)]
# Adding a journal manually because it fits, but is not selected by above method.
disc_w_journal[Title == "FINANCIAL ANALYSTS JOURNAL", Finance := TRUE]

disc_w_journal[is.na(Finance), Finance := FALSE]

# Saving the journal ids in the finance category
# (either in discipline economics or management)
disc_w_journal[Finance ==TRUE,list(Code_Discipline,journal_ID)] %>%
  write.csv(file = paste0(path_to_project_data, "finance_journals.csv"),
            row.names = FALSE)

```

*# The final list of journals in management that are put in the Finance category:*

```
disc_w_journal[Finance ==TRUE & Code_Discipline ==132,list(Title)]
```

```
## Title
## 1: ABACUS A JOURNAL OF ACCOUNTING FINANCE AND BUSINESS STUDIES
## 2: ACCOUNTING AND FINANCE
## 3: ASIA PACIFIC JOURNAL OF FINANCIAL STUDIES
## 4: CFO THE MAGAZINE FOR SENIOR FINANCIAL EXECUTIVES
## 5: COMPTABILITE CONTROLE AUDIT
## 6: CORPORATE GOVERNANCE AN INTERNATIONAL REVIEW
## 7: EUROPEAN FINANCIAL MANAGEMENT
## 8: EUROPEAN JOURNAL OF FINANCE
## 9: FAMILY BUSINESS REVIEW
## 10: FINANCE A UVER CZECH JOURNAL OF ECONOMICS AND FINANCE
## 11: FINANCE AND DEVELOPMENT
## 12: FINANCE AND STOCHASTICS
## 13: FINANCE AND TRADE REVIEW
## 14: FINANCE RESEARCH LETTERS
## 15: FINANCIAL ANALYSTS JOURNAL
## 16: FINANCIAL EXECUTIVE
## 17: FINANCIAL MANAGEMENT
## 18: FINANZARCHIV
## 19: INSURANCE MATHEMATICS AND ECONOMICS
## 20: INTERNATIONAL JOURNAL OF FINANCE AND ECONOMICS
## 21: INTERNATIONAL JOURNAL OF HEALTH CARE FINANCE AND ECONOMICS
## 22: INTERNATIONAL REVIEW OF FINANCE
## 23: INTERNATIONAL REVIEW OF FINANCIAL ANALYSIS
## 24: INVESTMENT ANALYSTS JOURNAL
## 25: JASSA THE FINSIA JOURNAL OF APPLIED FINANCE
## 26: JOURNAL OF BANKING AND FINANCE
## 27: JOURNAL OF BEHAVIORAL FINANCE
## 28: JOURNAL OF BUSINESS FINANCE
## 29: JOURNAL OF COMPUTATIONAL FINANCE
## 30: JOURNAL OF CORPORATE FINANCE
## 31: JOURNAL OF CREDIT RISK
## 32: JOURNAL OF DERIVATIVES
## 33: JOURNAL OF FINANCE
## 34: JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS
## 35: JOURNAL OF FINANCIAL ECONOMICS
## 36: JOURNAL OF FINANCIAL INTERMEDIATION
## 37: JOURNAL OF FINANCIAL MARKETS
## 38: JOURNAL OF FINANCIAL RESEARCH
## 39: JOURNAL OF FINANCIAL SERVICES RESEARCH
## 40: JOURNAL OF FUTURES MARKETS
## 41: JOURNAL OF INTERNATIONAL FINANCIAL MANAGEMENT AND ACCOUNTING
## 42: JOURNAL OF PORTFOLIO MANAGEMENT
## 43: JOURNAL OF RISK AND INSURANCE
## 44: JOURNAL OF RISK MODEL VALIDATION
## 45: MANAGERIAL FINANCE
## 46: MATHEMATICAL FINANCE
## 47: MATHEMATICS AND FINANCIAL ECONOMICS
## 48: PACIFIC BASIN FINANCE JOURNAL
## 49: PROCEEDINGS OF THE INSTITUTION OF MECHANICAL ENGINEERS PART 0 JOURNAL OF RISK AND RELIABILITY
```

```
## 50: QUANTITATIVE FINANCE
## 51: QUARTERLY REVIEW OF ECONOMICS AND FINANCE
## 52: REVIEW OF DERIVATIVES RESEARCH
## 53: REVIEW OF FINANCIAL STUDIES
## 54: REVISTA ESPANOLA DE FINANCIACION Y CONTABILIDAD SPANISH JOURNAL OF FINANCE AND ACCOUNTING
## 55: SIAM JOURNAL ON FINANCIAL MATHEMATICS
## 56: SPANISH JOURNAL OF FINANCE AND ACCOUNTING REVISTA ESPANOLA DE FINANCIACION Y CONTABILIDA
## 57: UNIVERSIA BUSINESS REVIEW
## Title
```

```
rm(disc_w_journal,list_j_CNRS_finance,list_j_scimago)
```

Angrist and colleagues had a list of finance journals. We load it and compare it to our list. First, we identify the journals that are in our list for finance and are also in Angrist's list:

```
# Loading files
load(paste0(path_to_project_data, "dt_angrist_journals.RData" ))
finance_journals <- read.csv(
  paste0(path_to_project_data, "finance_journals.csv")) %>% data.table()

dt_angrist_journals[Code_Revue %in% finance_journals$journal_ID,
  list(Code_Revue, Code_Discipline, Revue,
  angrist_disc)]
```

##	Code_Revue	Code_Discipline	Revue
## 1:	5589	132	FINANCIAL ANALYSTS JOURNAL
## 2:	5590	132	FINANCIAL MANAGEMENT
## 3:	8180	132	JOURNAL OF BANKING & FINANCE
## 4:	8495	132	JOURNAL OF CORPORATE FINANCE
## 5:	8788	132	JOURNAL OF FINANCE
## 6:	8789	132	JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS
## 7:	8791	132	JOURNAL OF FINANCIAL ECONOMICS
## 8:	8792	132	JOURNAL OF FINANCIAL INTERMEDIATION
## 9:	8793	132	JOURNAL OF FINANCIAL MARKETS
## 10:	8794	132	JOURNAL OF FINANCIAL RESEARCH
## 11:	8795	132	JOURNAL OF FINANCIAL SERVICES RESEARCH
## 12:	8845	132	JOURNAL OF FUTURES MARKETS
## 13:	9089	119	JOURNAL OF INTERNATIONAL MONEY AND FINANCE
## 14:	9680	132	JOURNAL OF PORTFOLIO MANAGEMENT
## 15:	9805	132	JOURNAL OF RISK AND INSURANCE
## 16:	9806	119	JOURNAL OF RISK AND UNCERTAINTY
## 17:	11065	132	MATHEMATICAL FINANCE
## 18:	11744	119	NATIONAL TAX JOURNAL
## 19:	13997	119	REVIEW OF FINANCE
## 20:	13998	132	REVIEW OF FINANCIAL STUDIES
## 21:	19155	119	ANNUAL REVIEW OF FINANCIAL ECONOMICS
##	Code_Revue	Code_Discipline	Revue
##	angrist_disc		
## 1:	Finance		
## 2:	Finance		
## 3:	Finance		
## 4:	Finance		
## 5:	Finance		
## 6:	Finance		
## 7:	Finance		

```
## 8:      Finance
## 9:      Finance
## 10:     Finance
## 11:     Finance
## 12:     Finance
## 13:     Finance
## 14:     Finance
## 15:     Finance
## 16:     Economics
## 17:     Finance
## 18:     Economics
## 19:     Finance
## 20:     Finance
## 21:     Finance
##      angrist_disc
```

```
rm(dt_angrist_journals,finance_journals)
```

We see that only one journal in this intersection is not identified by them as being in Finance: ‘Journal of Risk and Uncertainty’. Given that it falls in the code discipline of economics (119), this discrepancy has no impact on our results.

The second table lists the journals that are identified as finance by them, but not by us.

```
# loading files
load(paste0(path_to_project_data, "dt_angrist_journals.RData" ))
finance_journals <- read.csv(
  paste0(path_to_project_data, "finance_journals.csv")) %>% data.table()

dt_angrist_journals[angrist_disc == "Finance" &
  (!Code_Revue %in% finance_journals$journal_ID),
  list(Code_Revue, Code_Discipline, Revue,
  angrist_disc)]
```

```
##      Code_Revue Code_Discipline                               Revue
## 1:           8255             132                JOURNAL OF BUSINESS
## 2:          13715             132  QUARTERLY REVIEW OF ECONOMICS AND BUSINESS
## 3:          18232             119    ECONOMIC AND BUSINESS BULLETIN
##      angrist_disc
## 1:      Finance
## 2:      Finance
## 3:      Finance
```

```
rm(dt_angrist_journals,finance_journals)
```

There is no reason to think that their classification is better than ours in sorting these three journals.

Now, we start the investigation of the interface between economics journals, finance journals and other management journals. Which journals in the general category “Management” are most cited by economics?

```
core_SSH <- c("Psychology", "Economics",
  "Political Science &\nPublic Administration",
  "Sociology", "Management",
  "Anthropology &\n Archaeology")

dt_citing_cited <- readRDS(paste0(path_to_project_data,
  "dt_citing_cited_all.rds"))
```

```

# Restricting to econ and management
d_id_e_m <- c(119,132)

dt_citing_cited <- dt_citing_cited[citing_discipline %in% d_id_e_m &
  cited_discipline %in% d_id_e_m]

econ_cit_to_management <- dt_citing_cited[citing_discipline ==119 &
  cited_discipline ==132,.N,
  by=cited_code_revue]

# Fetching our list of journals
ESH <- dbConnect(MySQL(), user=usr, password=pswd, dbname='OST_Expanded_SciHum',
  host='127.0.0.1')
disc_w_journal <- dbGetQuery(ESH,
  "SELECT Code_Revue as journal_ID, Code_Discipline,
  Revue as Title
  FROM OST_Expanded_SciHum.Revues;") %>% data.table
disc_w_journal$Title <- strsplit(disc_w_journal$Title ,"\r") %>% unlist() %>%
  clean_text()

econ_cit_to_management <- merge(econ_cit_to_management,
  disc_w_journal[,list(journal_ID,Title)],
  by.x="cited_code_revue", by.y= "journal_ID")
finance_journals <- read.csv(paste0(path_to_project_data,
  "finance_journals.csv")) %>% data.table()

econ_cit_to_management[cited_code_revue %in% finance_journals$journal_ID,Finance:= TRUE]
econ_cit_to_management[is.na(Finance),Finance:=FALSE]
econ_cit_to_management[order(-N),list(`Journal ID`=cited_code_revue,
  Finance,citations = N, Title)]

```

```

##      Finance citations      Title
##  1:   TRUE      86293  JOURNAL OF FINANCE
##  2:   TRUE     54770  JOURNAL OF FINANCIAL ECONOMICS
##  3:  FALSE     27706  MANAGEMENT SCIENCE
##  4:   TRUE     27225  REVIEW OF FINANCIAL STUDIES
##  5:   TRUE     26448  JOURNAL OF BANKING AND FINANCE
## ---
## 420: FALSE         1  BRITISH JOURNAL OF MARKETING
## 421: FALSE         1      COST AND MANAGEMENT
## 422: FALSE         1  PUBLIC PERSONNEL REVIEW
## 423: FALSE         1  WHARTON QUARTERLY
## 424:  TRUE         1  COMPTABILITE CONTROLE AUDIT

```

```
rm(dt_citing_cited)
```

We see that the *Journal of Finance* is significantly more cited than the others and that there are 4 finance journals in the top 5. Yet, many other management journals have high citation counts.

Next, we can reproduce the graphs of COC for economics.

The first one is simply presenting the total COC of economics if finance is considered ‘outside’ or ‘inside’ economics.

```

dt_citing_cited <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))
dt_citing_cited <- dt_citing_cited[between(Year, 1950, 2018)]
# restricting to citing journals in economics
dt_citing_cited <- dt_citing_cited[citing_discipline == 119]

# Loading the list of finance journals
finance_journals <- read.csv(paste0(path_to_project_data,
                                   "finance_journals.csv")) %>% data.table()
# Creating a new code for discipline cited for finance
fin_disc <- data.table(Code_Discipline=200, discipline = "Finance")
disc_info <- rbind(discipline_info[,list(Code_Discipline, discipline)], fin_disc)

# relabeling finance journals with the new code discipline
dt_citing_cited[cited_code_revue %in% finance_journals[Code_Discipline == 132, journal_ID],
                cited_discipline:= fin_disc$Code_Discipline]

# Naming disciplines
dt_citing_cited <- merge(dt_citing_cited, disc_info[,list(Code_Discipline, discipline)],
                        by.x="cited_discipline", by.y="Code_Discipline", all.x=T)

# Citations per discipline x year
setkey(dt_citing_cited, Year, discipline)
citing_cited_econ <- dt_citing_cited[!is.na(discipline), list(Citations = .N),
                                   by=key(dt_citing_cited)]

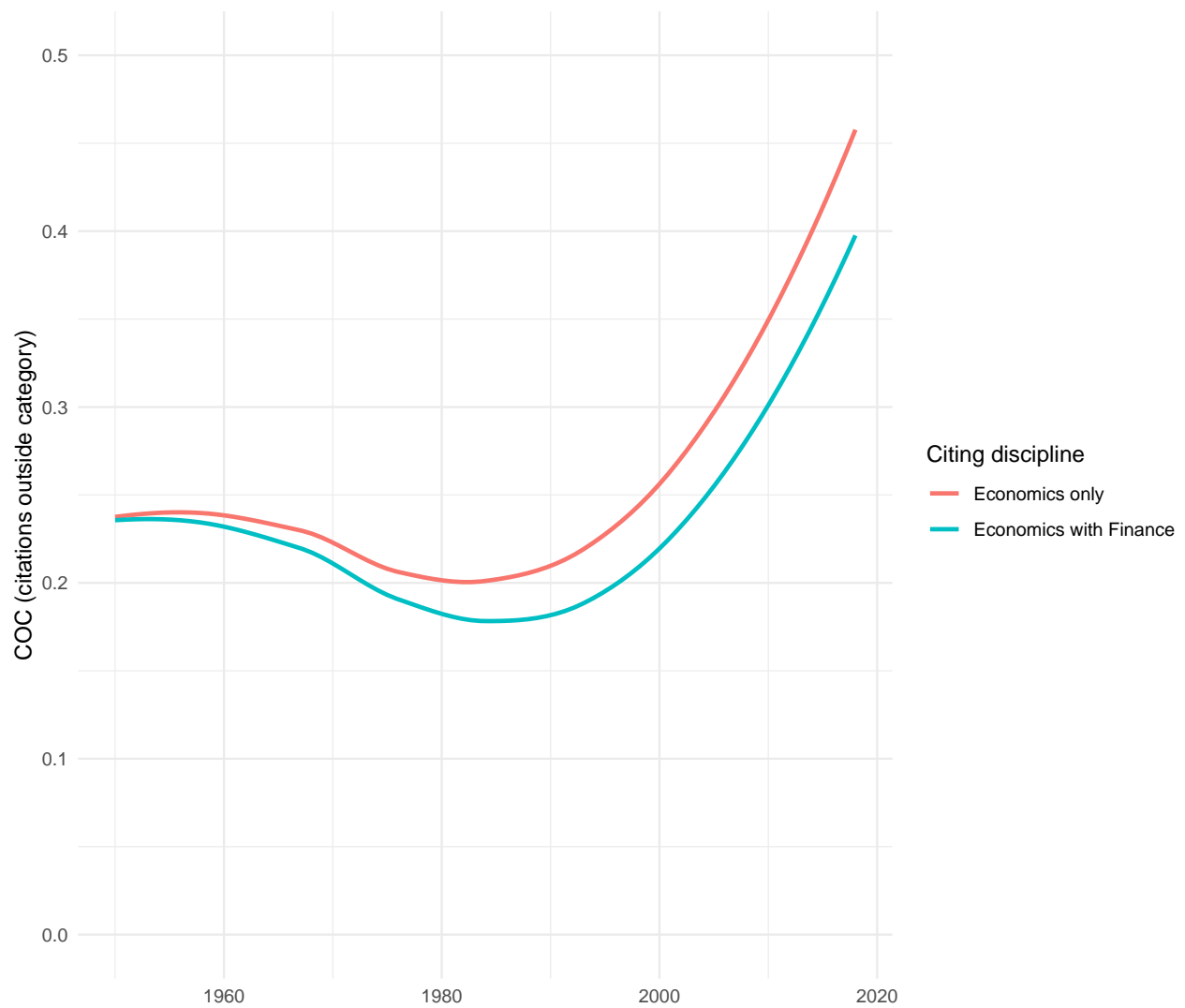
interdisc_econ_and_fin <- rbind(
  citing_cited_econ[,list(`Citing discipline`= "Economics only",
                        Proportion= 1- .SD[discipline == "Economics", Citations]/
                        sum(Citations)
                        ), by=Year],
  citing_cited_econ[,list(`Citing discipline`= "Economics with Finance",
                        Proportion= 1- .SD[discipline %in% c("Economics", "Finance"),
                        sum(Citations)]/sum(Citations)
                        ), by=Year]
)

ggplot(interdisc_econ_and_fin, aes(x= Year, y= Proportion, color = `Citing discipline`)) +
  geom_smooth(se=F, method="loess") + xlab("") + ylab("COC (citations outside category)") +
  ylim(0, .5) + theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'

```





```
rm(dt_citing_cited,citing_cited_econ, interdisc_econ_and_fin)
```

We see that the interdisciplinary turn remains important even if finance is taken to be integrated to economics.

Our next and last step is to disaggregate the COC of economics as before, but to split management into ‘Finance’ and ‘Management (other)’.

```
the_disc <- "Economics"
the_varia_SSH <- "Others SSH"
the_varia_NSE <- "Others NSE"
SSH <- "Social Sciences and Humanities"
NSE <- "Natural Sciences and Engineering"

#Data needed
dt_citing_cited <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))
dt_citing_cited <- dt_citing_cited[between(Year, 1950, 2019),]

# Loading the list of finance journals
finance_journals <- read.csv(paste0(path_to_project_data,
                                   "finance_journals.csv")) %>% data.table()

# Creating a new code for discipline cited for finance
```

```

fin_disc <- data.table(EGrande_Discipline= SSH, Code_Discipline=200,
                      discipline = "Finance")
disc_info <- rbind(discipline_info[,list(EGrande_Discipline,
                                         Code_Discipline,discipline)],fin_disc)
disc_info[Code_Discipline == 132,discipline:= "Management (other)"]
# relabeling finance journals with the new code discipline
dt_citing_cited[cited_code_revue %in%
                 finance_journals[Code_Discipline ==132, journal_ID],
                 cited_discipline:= fin_disc$Code_Discipline]
dt_citing_cited[citing_code_revue %in%
                 finance_journals[Code_Discipline ==132, journal_ID],
                 citing_discipline:= fin_disc$Code_Discipline]

#Summing how many citation goes from every dicipline to every other discipline per year
citing_cited_disc <- dt_citing_cited[, list(Citations =.N),
                                       by=c("Year", "citing_discipline",
                                             "cited_discipline")]
#Merging with dicipline_info to get cited discipline names
citing_cited_disc <- merge(citing_cited_disc,
                           disc_info[,list(EGrande_Discipline, Code_Discipline,
                                             Discipline_cited = discipline)],
                           by.x="cited_discipline", by.y="Code_Discipline", all.x=T)

#Merging with dicipline_info to get citing discipline names
citing_cited_disc <- merge(citing_cited_disc,
                           disc_info[,list(Code_Discipline,
                                             Discipline_citing = discipline)],
                           by.x="citing_discipline", by.y="Code_Discipline", all.x=T)

#Getting references for Economics only
cit_disc <- citing_cited_disc[Discipline_citing == the_disc & !is.na(cited_discipline)]

#Summing citations per dicipline cited
cit_disc <- cit_disc[,list(Citations= sum(Citations)),
                     by=list(Year,Discipline_citing,Discipline_cited,
                              EGrande_Discipline)]

setkey(cit_disc,Year)

#Calculating proportion and taking relevant columns
cit_disc <- cit_disc[, list(EGrande_Discipline, Discipline_cited,
                           Proportion = Citations/sum(Citations)),by=Year]

# Finding out the most important disciplines:
n_disc_now <- 8
imp_disc <- cit_disc[Discipline_cited != the_disc, list(highest_prop = max(Proportion)),
                    by=Discipline_cited][order(-highest_prop)][1:n_disc_now,
                    Discipline_cited]
dt_p <- cit_disc[Discipline_cited %in% imp_disc, list(Year, Discipline_cited,
                                                    Proportion)]

#Adding a all_others column for plotting purposes

```

```

dt_p$all_others <- c("Cited Disciplines")

#Agregating all other cited discipline and dividing it in Other SSH and Other NSE
dt_p_all_others <- cit_disc[!Discipline_cited %in% c(imp_disc,the_disc) &
  EGrande_Discipline == SSH,
  list(Discipline_cited=the_varia_SSH,
    Proportion= sum(Proportion)),by=Year]
dt_p_all_others <- rbind(dt_p_all_others,
  cit_disc[!Discipline_cited %in% c(imp_disc,the_disc) &
    EGrande_Discipline == NSE,
    list(Discipline_cited=the_varia_NSE,
      Proportion= sum(Proportion)),by=Year])

dt_p_all_others[, all_others := Discipline_cited]

#Merging important discipline with all others
dt_p <-rbind(dt_p, dt_p_all_others))

#Colors for every disciplines.
colors_others<- c("Earth and Space" = alpha("brown", 0.7),
  "Sociology" = "#FF0000",
  "Mathematics" = alpha("darkgreen", 0.6),
  "Biomedical Research" = alpha("#00e5e5", 0.7),
  "Management (other)" = "purple",
  "Biology" = alpha("green", 0.7),
  "Other Engineering \n and Technology" = alpha("#E59400", 0.7),
  "Psychology" = "#0000FF",
  "Finance" = "orange")

#Linetypes for every disciplines.
linetypes_others <- c("Earth and Space" = "solid",
  "Sociology" = "dotted",
  "Mathematics" = "solid",
  "Biomedical Research" = "solid",
  "Management (other)" = "dotdash",
  "Biology" = "solid",
  "Other Engineering \n and Technology" = "solid",
  "Psychology" = "longdash",
  "Finance" = "dashed")

#Ordering for plotting purposes
order_disc <- unique(dt_p[, Prop := predict(loess(Proportion~Year), Year),
  by = Discipline_cited][Year == 2018,][
  order(-Prop),Discipline_cited])
dt_p$Discipline_cited <- factor(dt_p$Discipline_cited, levels = order_disc)

ggplot(dt_p[all_others == "Cited Disciplines"]) +
  geom_smooth(se=F,method = 'loess',aes(x=Year, y=Prop,col = Discipline_cited,
    linetype = Discipline_cited)) +
  scale_color_manual(name = "Disciplines", values = colors_others) +
  scale_linetype_manual(name = "Disciplines", values = linetypes_others) +

```

```

guides(linetype=guide_legend(keywidth = 1.5, keyheight = 1,byrow=TRUE),
       color=guide_legend(keywidth = 1.5, keyheight = 1,byrow=TRUE)) +
new_scale_linetype() + # geoms below can use another linetype scale!
geom_smooth(dt_p[all_others != "Cited Disciplines"],
            mapping = aes(x=Year, y = Prop,#, group = Discipline_for_plot
                          linetype = all_others),
                          color = "gray",
                          se=F, method = 'loess') +
scale_linetype_manual(name="Other Disciplines",
                      values = c("Others NSE" = "longdash",
                                  "Others SSH" = "dashed")) +
#theme(legend.box = "horizontal") +
ylab("COC (citations outside category)") + theme_minimal() +
theme(legend.position="bottom", legend.box = "vertical",
      axis.text=element_text(size=13),
      axis.title=element_text(size=14),
      legend.text=element_text(size=13),
      legend.title=element_text(size=14)
      ) +
annotate( "text",x=max(dt_p$Year),y=-Inf,hjust=.95,vjust=-0.5,
          label="Curves smoothed by local polynomial regression",
          ize=4,color="darkgrey" )

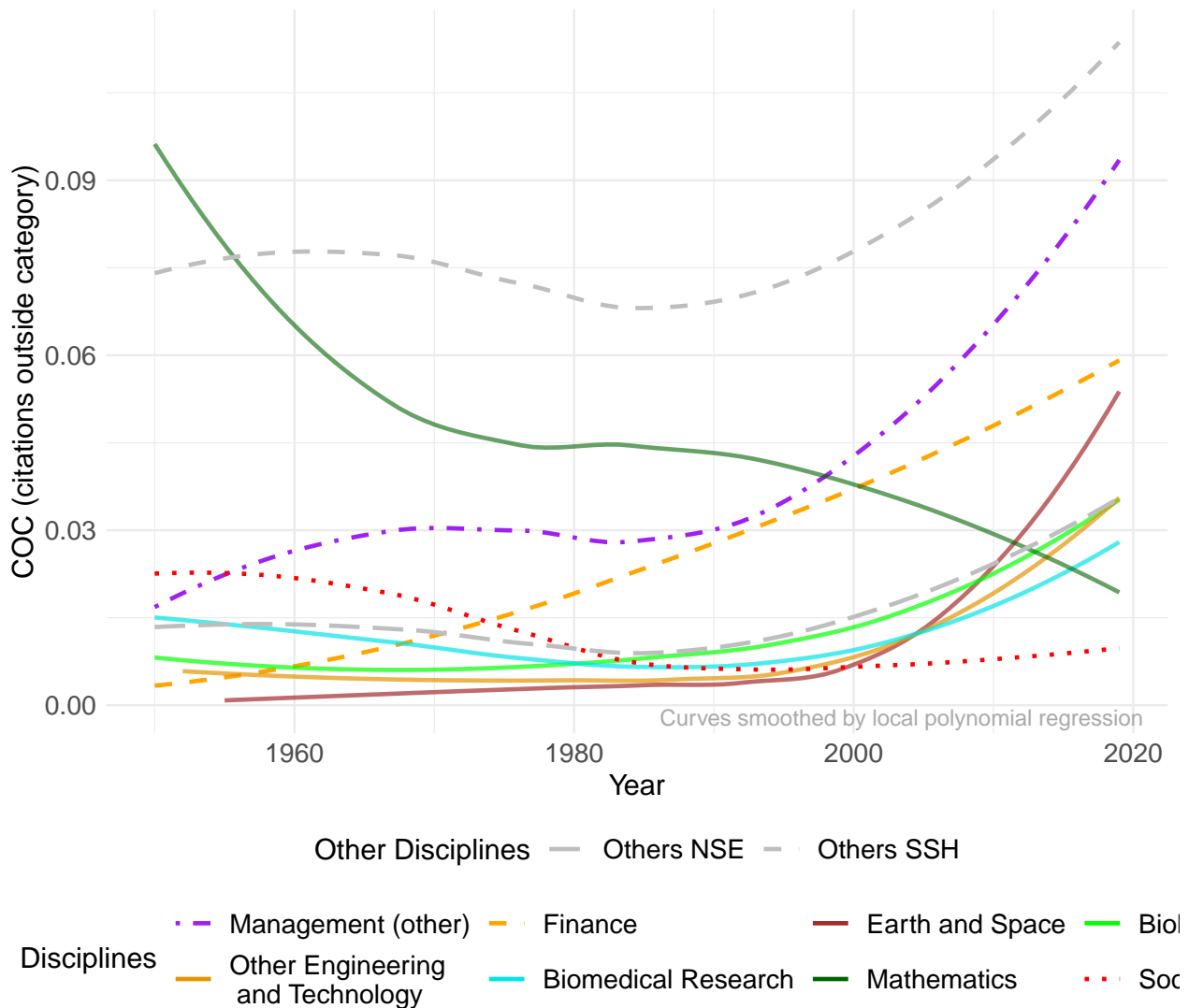
```

```

## Warning in annotate("text", x = max(dt_p$Year), y = -Inf, hjust = 0.95, :
## Ignoring unknown parameters: `ize`

## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'

```



```
ggsave( paste0("figures/", "fig13.png"), width = 16, height = 9, bg = "white",
  scale = .7)
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

```
rm(dt_citing_cited); rm(citing_cited_disc)
```

We see that the proportion of references to finance grows with time, but the proportion of references to the rest of management grows at an even stronger pace since the 1990s. In sum, the rising interdisciplinarity of economics is not the story of the rising connection to finance.

### 5.3 Weak ties of economics

The goal now is to have a clearer view of what is happening at the bottom of the previous figures. We are thus reproducing the analysis of economics' COC, but excluding management and mathematics. We are also adding some disciplines to have 10 in total.

```
#Data needed
dt_citing_cited <- readRDS(paste0(path_to_project_data, "dt_citing_cited_all.rds"))
```

```

dt_citing_cited <- dt_citing_cited[between(Year, 1950, 2019),]
#Summing how many citation goes from every dicipline to every other discipline per year
citing_cited_disc <- dt_citing_cited[, list(Citations =.N),
                                     by=c("Year", "citing_discipline",
                                           "cited_discipline")]
#Merging with dicipline_info to get cited discipline names
citing_cited_disc <- merge(citing_cited_disc,
                          discipline_info[,list(EGrande_Discipline, Code_Discipline,
                                                Discipline_cited = discipline)],
                          by.x="cited_discipline", by.y="Code_Discipline", all.x=T)
#Merging with dicipline_info to get citing discipline names
citing_cited_disc <- merge(citing_cited_disc,
                          discipline_info[,list(Code_Discipline,
                                                Discipline_citing = discipline)],
                          by.x="citing_discipline", by.y="Code_Discipline", all.x=T)

rm(dt_citing_cited)

#The disciplines we will do the treatment on
the_disc <- "Economics"
n_disc <- 10
the_discs <- c("Earth and Space", "Biology", "Other Engineering \n and Technology",
              "Psychology", #"Health",
              "Other Social Sciences",
              "Political Science &\nPublic Administration",
              "Geography", "Sociology", "Biomedical Research",
              "Computers & \n Operations Research")

all_disc_colors <- c( "Earth and Space" = alpha("brown", 0.5),
                    "Sociology" = "#FF0000",
                    "Biomedical Research" = alpha("#00e5e5", 0.5),
                    "Biology" = alpha("green", 0.5),
                    "Other Engineering \n and Technology" = alpha("#E59400", 0.5),
                    "Health" = alpha("#C8A2C8", 0.5),
                    "Other Social Sciences" = alpha("#E59400", 0.7),
                    "Political Science &\nPublic Administration" = "#228B22",
                    "Geography" = alpha("#BABA45", 0.5),
                    "Computers & \n Operations Research" = "#A0A0A0",
                    "Psychology" = "#0000FF")

all_disc_linetypes <- c( "Earth and Space" = "solid",
                       "Sociology" = "dotted",
                       "Biomedical Research" = "solid",
                       "Biology" = "solid",
                       "Other Engineering \n and Technology" = "solid",
                       "Health" = "solid",
                       "Other Social Sciences" = "dotted",
                       "Political Science &\nPublic Administration" = "dashed",
                       "Geography" = "solid",
                       "Computers & \n Operations Research" = "solid",
                       "Psychology" = "longdash")

```

```

#Taking data for economic only
cit_disc <- citing_cited_disc[Discipline_citing == the_disc & !is.na(cited_discipline)]
#summing number of citation by dicipline cited
cit_disc <- cit_disc[,list(Citations= sum(Citations)),
                        by=list(Year,Discipline_citing,Discipline_cited,
                               EGrande_Discipline)]

setkey(cit_disc,Year)

#Calculating proportions
cit_disc <- cit_disc[, list(Discipline_cited, Proportion = Citations/sum(Citations)),
                        by=Year]

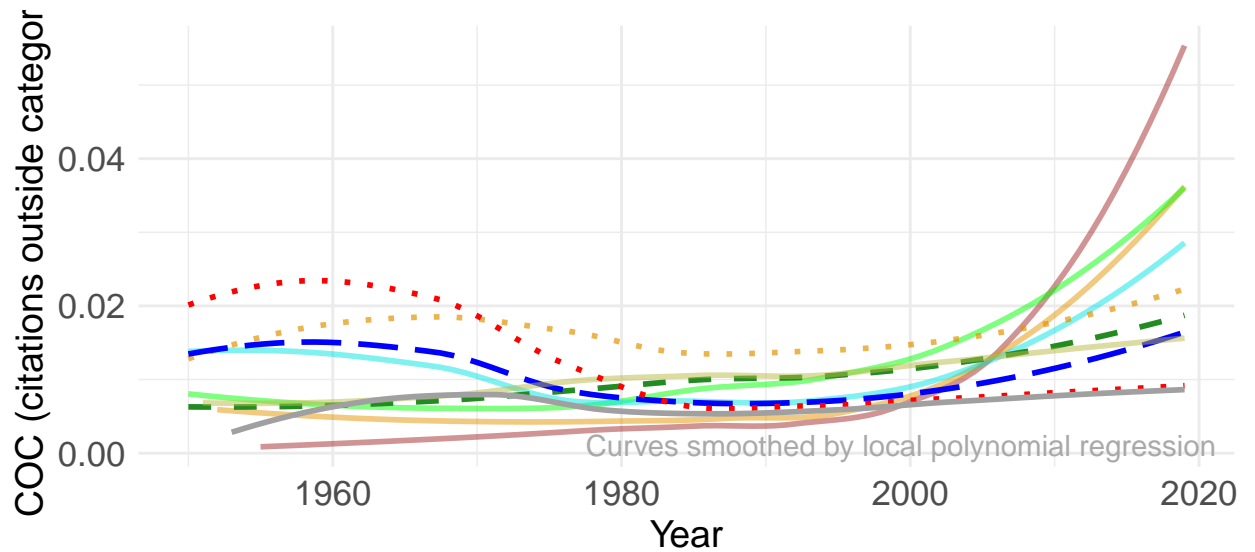
#Taking only disciplines we want to track
dt_p <- cit_disc[Discipline_cited %in% the_discs,
                list(Year, Discipline_cited, Proportion)]

#Ordering for plotting purposes
order_disc <- unique(dt_p[, Prop := predict(loess(Proportion~Year), Year),
                        by = Discipline_cited][Year ==
setkey(dt_p,Year,Discipline_cited)
#Factoring Discipline_cited assures we have a line for every discipline in our plot
dt_p$Discipline_cited <- factor(dt_p$Discipline_cited, levels = order_disc)
order_disc = order_disc[!order_disc %in% c("Others SSH", "Others NSE")]

#Plotting
ggplot() + geom_smooth(dt_p, mapping = aes(x=Year, y=Proportion, col = Discipline_cited,
                                           linetype = Discipline_cited), se=F,
                        method = 'loess') +
  scale_color_manual(name="Disciplines", values = all_disc_colors) +
  scale_linetype_manual(name="Disciplines", values = all_disc_linetypes) +
  #ggtitle(paste0("Extradisciplinary citations in ", the_disc,
  #              " without management or mathematics")) +
  guides(linetype=guide_legend(keywidth = 1, keyheight = 1,nrow=3,byrow=TRUE),
         color=guide_legend(keywidth = 1, keyheight = 1,nrow=3,byrow=TRUE)) +
  ylab("COC (citations outside category)") + theme_minimal() +
  theme(legend.position="bottom", legend.box = "vertical",
        axis.text=element_text(size=13),
        axis.title=element_text(size=14),
        legend.text=element_text(size=13),
        legend.title=element_text(size=14)) +
  annotate( "text",x=max(dt_p$Year),y=-Inf,hjust=.95,vjust=-0.5,
          label="Curves smoothed by local polynomial regression",
          size=4,color="darkgrey" )

## `geom_smooth()` using formula = 'y ~ x'

```



- Earth and Space
- Other Engineering and Technology
- Biology
- Bior
- Other Social Sciences
- Political Science & Public Administration
- Psychology
- Geoc
- Sociology
- Computers & Operations Research

```
ggsave( paste0("figures/", "fig14.png"), width = 16, height = 9, bg = "white",
        scale = .7)
```

```
## `geom_smooth()` using formula = 'y ~ x'
rm(citing_cited_disc);
```

### 5.3.1 Weighed citations

It is possible to produce the same figure but weighing citations based on the journal where they originate.

```
the_disc <- "Economics"
the_varia_SSH <- "Others SSH"
the_varia_NSE <- "Others NSE"
SSH <- "Social Sciences and Humanities"
NSE <- "Natural Sciences and Engineering"

dt_citing_cited_journals <- readRDS(paste0(path_to_project_data,
                                          "dt_citing_cited_journals.rds"))

load(paste0(path_to_project_data, "journal_impact_per_window.RData"))
setnames(dt_citing_cited_journals, "Year", "citing_Year")

# Focus on econ
cit_disc <- dt_citing_cited_journals[Discipline_citing == the_disc &
                                     !is.na(Discipline_cited),
                                     list(citing_code_revue, Discipline_citing,
```



```

Discipline_cited,
citing_Year, N )][
  ,list(Nb_cited =sum(N)),
  by = list(citing_code_revue,
            Discipline_cited,
            citing_Year) ]

# Getting the weights in there.
cit_disc <- merge(cit_disc,
  journal_impact_per_window[
    ,list(journal_ID,citing_Year,
          Journal_weight_in_disc =impact_weight )],
  by.x = c("citing_code_revue", "citing_Year"),
  by.y = c("journal_ID", "citing_Year"))

## Calculating proportions of extradisciplinary citations
#Proportion of extradisciplinary citations to various disciplines:
setkey(cit_disc,citing_Year)
cit_disc[,tot_weighted_cit:= sum(Nb_cited*Journal_weight_in_disc),by=citing_Year]
cit_disc <- cit_disc[, list(
  Proportion = sum(Nb_cited*Journal_weight_in_disc)/unique(tot_weighted_cit)
  ),
  by =.(Discipline_cited, citing_Year)]
setnames(cit_disc,"citing_Year","Year")

# Merging with discipline info to get EGrande_Disc
cit_disc <- merge(cit_disc,
  unique(discipline_info[,list(EGrande_Discipline,
                               Discipline_cited = discipline)]),
  by= "Discipline_cited", all.x=T)

#The disciplines we will do the treatment on
the_disc <- "Economics"
n_disc <- 10
the_discs <- c("Earth and Space", "Biology", "Other Engineering \n and Technology",
  "Psychology", #"Health",
  "Other Social Sciences",
  "Political Science &\nPublic Administration",
  "Geography", "Sociology",
  "Biomedical Research", "Computers & \n Operations Research")

all_disc_colors <- c( "Earth and Space" = alpha("brown", 0.5),
  "Sociology" = "#FF0000",
  "Biomedical Research" = alpha("#00e5e5", 0.5),
  "Biology" = alpha("green", 0.5),
  "Other Engineering \n and Technology" = alpha("#E59400", 0.5),
  #"Health" = alpha("#C8A2C8", 0.5),
  "Other Social Sciences" = alpha("#E59400", 0.7),
  "Political Science &\nPublic Administration" = "#228B22",
  "Geography" = alpha("#BABA45", 0.5),

```

```

    "Computers & \n Operations Research" = "#A0A0A0",
    "Psychology" = "#0000FF")

all_disc_linetypes <- c( "Earth and Space" = "solid",
    "Sociology" = "dotted",
    "Biomedical Research" = "solid",
    "Biology" = "solid",
    "Other Engineering \n and Technology" = "solid",
    # "Health" = "solid",
    "Other Social Sciences" = "dotted",
    "Political Science &\nPublic Administration" = "dashed",
    "Geography" = "solid",
    "Computers & \n Operations Research" = "solid",
    "Psychology" = "longdash")

#Taking only disciplines we want to track
dt_p <- cit_disc[Discipline_cited %in% the_discs, list(Year, Discipline_cited,
    Proportion)]

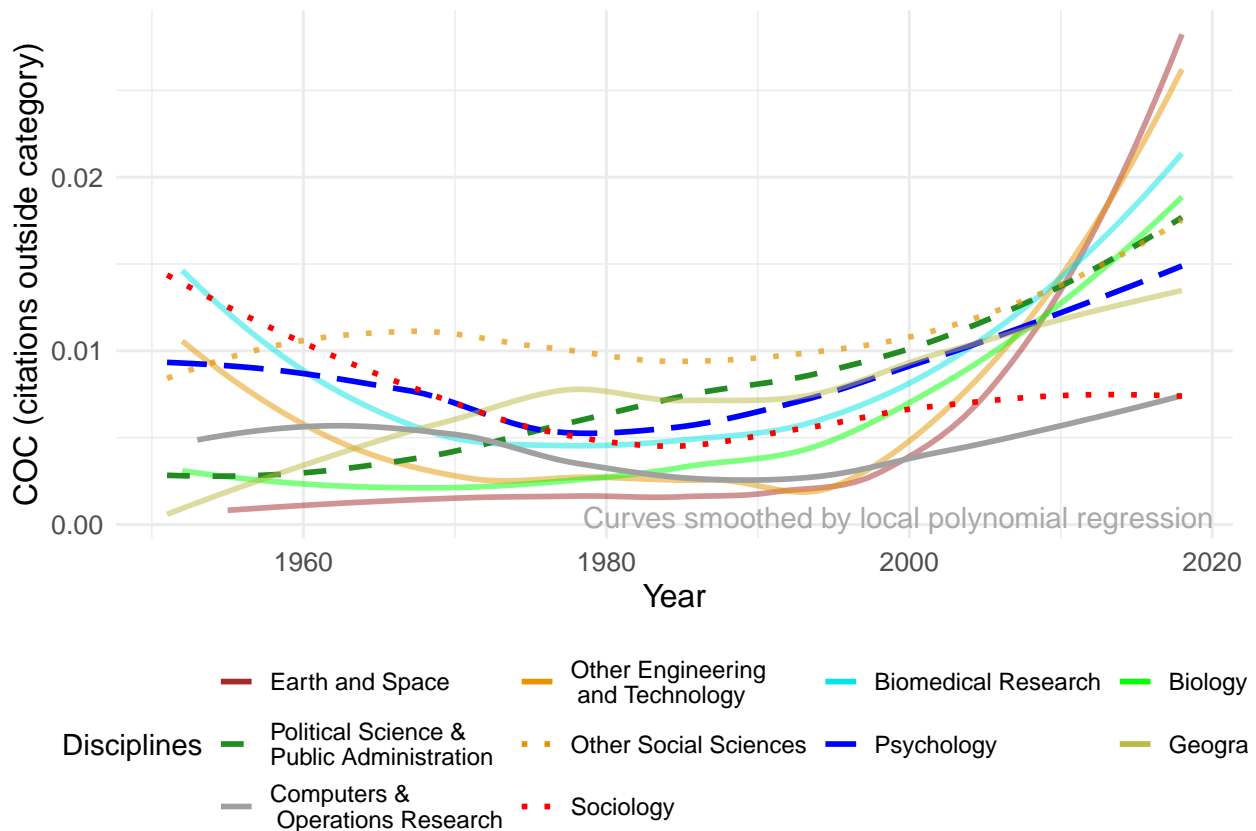
#Ordering for plotting purposes
order_disc <- unique(dt_p[, Prop := predict(loess(Proportion~Year), Year),
    by = Discipline_cited][Year ==
    max(Year),][
    order(-Prop),
    Discipline_cited])

setkey(dt_p,Year,Discipline_cited)
#Factoring Discipline_cited assures we have a line for every discipline in our plot
dt_p$Discipline_cited <- factor(dt_p$Discipline_cited, levels = order_disc)
order_disc = order_disc[!order_disc %in% c("Others SSH", "Others NSE")]

#Plotting
ggplot() + geom_smooth(dt_p, mapping = aes(x=Year, y=Proportion, col = Discipline_cited,
    linetype = Discipline_cited),
    se=F,method = 'loess') +
    scale_color_manual(name="Disciplines", values = all_disc_colors) +
    scale_linetype_manual(name="Disciplines", values = all_disc_linetypes) +
    #ggtitle(paste0("Extradisciplinary citations in ", the_disc,
    # " without management or mathematics")) +
    guides(linetype=guide_legend(keywidth = 1, keyheight = 1,nrow=3,byrow=TRUE),
    color=guide_legend(keywidth = 1, keyheight = 1,nrow=3,byrow=TRUE)) +
    ylab("COC (citations outside category)") +
    theme_minimal() + theme(legend.position="bottom", legend.box = "vertical",
    axis.text=element_text(size=10),
    axis.title=element_text(size=12#,face="bold"
    )) +
    annotate( "text",x=max(dt_p$Year),y=-Inf,hjust=.95,vjust=-0.5,
    label="Curves smoothed by local polynomial regression",
    size=4,color="darkgrey" )

## `geom_smooth()` using formula = 'y ~ x'

```



And again the results are similar to those with unweighed citations.

## 6 Comparison with the recent literature

We already reproduced above the main figure in Fourcade et al. We now turn to reproducing the analysis in Angrist et al.

In the following code we create a table containing the weight given to every journal each year using Angrist and al. methodology.

```
load(paste0(path_to_project_data, "dt_angrist_journals.RData" ))
dt_citing_cited_journals <- readRDS(paste0(path_to_project_data,
                                           "dt_citing_cited_journals.rds"))

# Creating a table for trunk journals. We have found the journal codes by hand.
# The trunk journals are found in Table A1 (Angrist and al., 2018, p. 63)
trunk_journals <- data.table()
trunk_journals$discipline <- unique(dt_angrist_journals$angrist_disc)
trunk_journals[discipline=="Sociology", Code_Revue := 936]
trunk_journals[discipline=="Economics", Code_Revue := 758]
trunk_journals[discipline=="Anthropology", Code_Revue := 736]
trunk_journals[discipline=="Political Science", Code_Revue := 926]
trunk_journals[discipline=="Psychology", Code_Revue := 13572]
trunk_journals <- rbind(trunk_journals, list("Psychology", 13573))
trunk_journals[discipline=="Management", Code_Revue := 31]
trunk_journals <- rbind(trunk_journals, list("Management", 28))
trunk_journals[discipline=="Accounting", Code_Revue := 39]
```

```

trunk_journals[discipline=="Marketing", Code_Revue := 9223]
trunk_journals[discipline=="Finance", Code_Revue := 8788]

#Renaming political science and Anthropology
dt_citing_cited_journals <- renaming_ssh Angrist(dt_citing_cited_journals)

#Renaming business disciplines
dt_citing_cited_journals <- renaming_business_Angrist(dt_citing_cited_journals,
                                                    dt_Angrist_journals)

##Journals' weight is equal to the proportion of cited articles by the trunk journal(s)
# going to that journal at a given year (this is Angrist and al. methodology)##
weight_Angrist <- dt_citing_cited_journals[!is.na(Discipline_cited) &
                                           citing_code_revue %in%
                                           trunk_journals$Code_Revue,
                                           list(citing_Year = Year, Discipline_citing,
                                                cited_code_revue, Discipline_cited, N)]

#Journal weight depends on citations within the discipline
weight_Angrist <- weight_Angrist[Discipline_citing == Discipline_cited]
#Counting number of citation every journal receives from trunk journals
weight_Angrist[,cited_journal_total := sum(N),
               by = .(citing_Year, Discipline_citing,
                     cited_code_revue, Discipline_cited)][, `:=`(N = NULL)]

#Getting rid of duplicates
weight_Angrist <- unique(weight_Angrist)
#Summing total number of citation by trunk journal
weight_Angrist[, discipline_total := sum(cited_journal_total),
               by = .(citing_Year, Discipline_citing)]

#Calculating proportions
weight_Angrist[, discipline_proportion := cited_journal_total/discipline_total,
               by = citing_Year][, `:=`(cited_journal_total = NULL,
                                         discipline_total = NULL,
                                         Discipline_cited = NULL)]

save(weight_Angrist, file = paste0(path_to_project_data, "weight_Angrist.RData"))
rm(dt_Angrist_journals, dt_citing_cited_journals, weight_Angrist, trunk_journals)

```

One question is the extent to which top journals are weighed more relative to the rest with their measure of 'journal importance.' We simply take their measure of journal importance for economics as reported in their Table W1 and check how many journals at the bottom of their list weigh as much as the unique trunk journal (the *American Economic Review*)

```

## Checking the skewed weights of Angrist et al. ####

# Loading the names of journals
eco_ref <- read.csv(paste0(path_to_project_data, "economics.csv"),
                  header = F,col.names = "Journal") %>% data.table()
eco_ref$Journal <- trimws(eco_ref$Journal) # removing leading and trailing spaces

# Manually inputting the weights (taken from table W1 in Angrist et al)
eco_ref$importance <- c(.261,.127,.086,.077,.046,.033,.031,.03,.022,.022,.022,0.019,
                      0.019, .018,.014,.014,.013,.011,.011,rep(.01,3),rep(.009,3),
                      .008,.008,rep(.007,3),rep(.006,7),rep(.005,6),rep(.004,11),

```

```

rep(.003,7),rep(.002,7),.001)

# AER only:
imp_AER <- eco_ref[,.SD[1,sum(importance)]/sum(importance)] %>% round(.,2)

# The bottom with the same share as AER.
lower_set <- 19:nrow(eco_ref)
imp_bottom <- eco_ref[,.SD[lower_set,sum(importance)]/sum(importance)] %>% round(.,2)
nb_bottom <- lower_set %>% length()

rm(eco_ref,lower_set)

# And loading two types of weights for further investigation
load(paste0(path_to_project_data, "journal_impact_per_window.RData"))
load(paste0(path_to_project_data, "weight Angrist.RData"))

```

The share of AER is 0.23, which corresponds to the share of 0.23 of the bottom 51 journals (among 69 in total).

Another way to see the extent to which Angrist et al's procedure overestimates the importance of the *American Economic Review* is to compare the average weight of this journal according to their procedure – i.e., 0.249 – with its average weight according to our own (more sensible) impact factor: 0.047. They thus overestimate the AER's importance by: 524%!

Now, we reproduce their Figure 1:

```

#Loading data
dt_citing_cited_journals <- readRDS(paste0(path_to_project_data,
                                           "dt_citing_cited_journals.rds" ))
load(paste0(path_to_project_data, "weight Angrist.RData"))
load(paste0(path_to_project_data, "dt Angrist_journals.RData" ))

#Renaming discipline to match Angrist's category names
dt_citing_cited_journals <- renaming_ssh_Angrist(dt_citing_cited_journals)
dt_citing_cited_journals <- renaming_business_Angrist(dt_citing_cited_journals,
                                                    dt_Angrist_journals)
setnames(dt_citing_cited_journals,"Year","citing_Year")

#Lists of disciplines per graph panel
core_AngRist <- c("Psychology","Economics",
                 "Political Science",
                 "Sociology", "Anthropology")
business_AngRist <- c("Marketing","Management",
                    "Finance", "Accounting")
other_disciplines <- c(33:50, 52:59, 61:66, 77, 87:91, 92:100, 113)

# Taking relevant columns and summing number of citation going from every discipline
# to every other (including itself)
dt_p <- dt_citing_cited_journals[!is.na(Discipline_cited) &
                                Discipline_citing %in% core_AngRist,
                                list(citing_code_revue, Discipline_citing,
                                     Discipline_cited, cited_discipline,
                                     citing_Year , N )]

```

```

dt_p <- dt_p[cited_discipline %in% other_disciplines,
             Discipline_cited := "Other_discipline"][,sum(N),
                                                    by = list(citing_code_revue,
                                                            Discipline_citing,
                                                            Discipline_cited,
                                                            citing_Year) ]

setnames(dt_p, "V1", "Nb_cited")

#Merging Angrist's weights
dt_p <- merge(dt_p, weight_angrist,
             by.x = c("Discipline_citing", "citing_code_revue", "citing_Year"),
             by.y = c("Discipline_citing", "cited_code_revue", "citing_Year"))
setnames(dt_p, "discipline_proportion", "Journal_weight_in_disc")

## Calculating proportions of extradisciplinary citations
dt_p[, Total_journal := sum(Nb_cited), by = .(citing_Year, citing_code_revue)][
  , Proportion := Nb_cited / Total_journal][, `:=`(Nb_cited = NULL,
                                                    Total_journal = NULL)]

## Calculating weighted proportions
dt_p[, weighted_proportion := Journal_weight_in_disc * Proportion][
  , `:=`(Journal_weight_in_disc = NULL, Proportion = NULL)]

#This chunk of code makes it possible to have three graphs next to each other in ggplot
dt_p[Discipline_cited %in% core_angrist, Graph := "1. Social Sciences"]
dt_p[Discipline_cited %in% business_angrist, Graph := "2. Business Disciplines"]
dt_p[Discipline_cited == "Other_discipline", Graph := "3. Other Disciplines"]

#Getting rid of discipline that are not tracked by Angrist
dt_p <- dt_p[!is.na(Graph) & Discipline_citing != Discipline_cited]

## Summing weighted proportions
dt_plot_angrist <- dt_p[, sum(weighted_proportion),
                       by = .(Discipline_citing, Graph, citing_Year)]
setnames(dt_plot_angrist, "V1", "Proportion")

#Ordering discipline manually so they appear in that order in the graph's legend.
order_disc <- c("Economics", "Political Science", "Sociology",
               "Psychology", "Anthropology")
dt_plot_angrist$Discipline_citing <- factor(dt_plot_angrist$Discipline_citing,
                                           levels = order_disc)
dt_plot_angrist <- dt_plot_angrist[between(citing_Year, 1970, 2015)]

#Assigning colors and linetype to disciplines to make our graph look like Angrist's
colors_angrist <- c("Psychology" = "blue", "Economics" = "black", "Sociology" = "red",
                  "Anthropology" = "#CCCC00", "Political Science" = "#228B22")
linetypes_angrist <- c("Psychology" = "longdash", "Economics" = "solid",
                     "Sociology" = "dotted", "Anthropology" = "solid",
                     "Political Science" = "dashed")

#Plotting
ggplot(dt_plot_angrist, aes(x=citing_Year, y=Proportion, color=Discipline_citing)) +

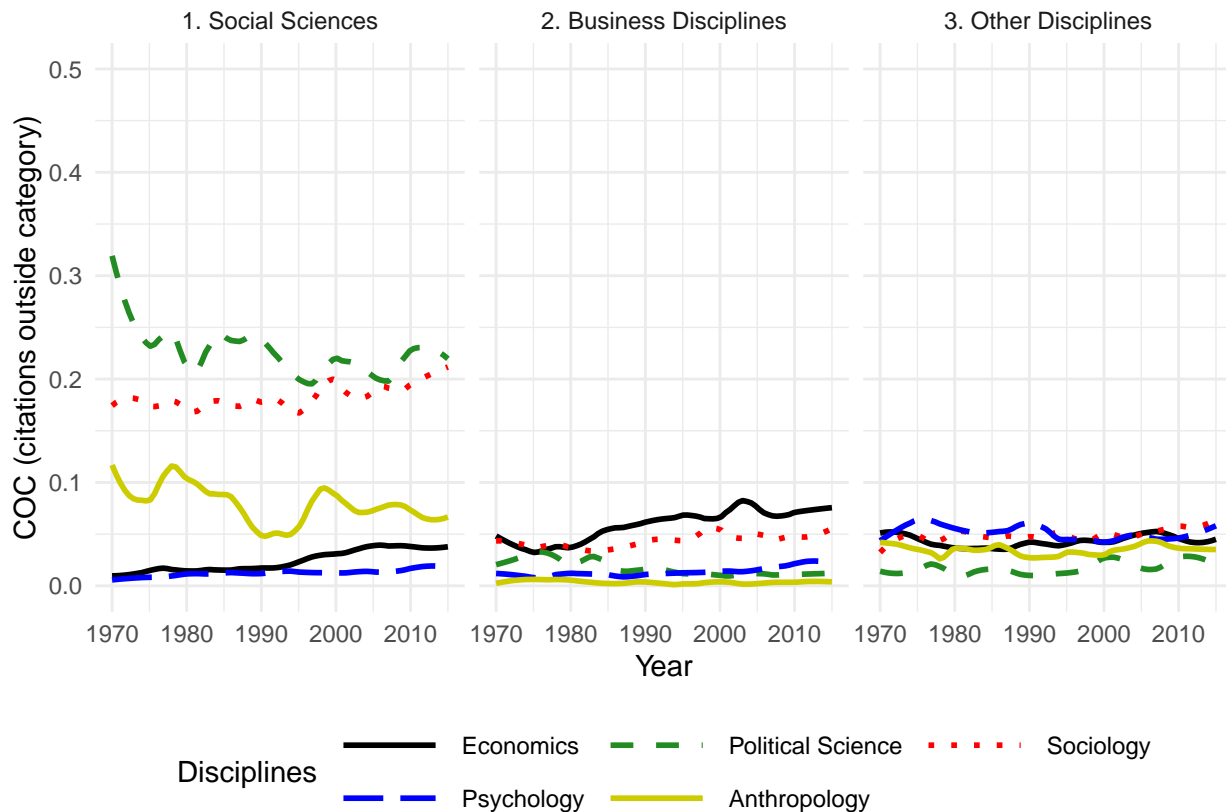
```

```

geom_smooth(se = F, span = 0.25, aes(linetype = Discipline_citing)) +
labs(x="Year",y="COC (citations outside category)",col="Discipline") +
#ggtitle("Social Science Insularity (Fig. 1 reproduction)") +
facet_wrap(~ Graph, scales = "free_x") + ylim(0, 0.5) + theme_minimal() +
theme(legend.position="bottom") +
scale_color_manual(name = "Disciplines", values = colors_angrist) +
scale_linetype_manual(name = "Disciplines", values = linetypes_angrist) +
guides(linetype=guide_legend(keywidth = 3, keyheight = 1,nrow=2,byrow=TRUE),
color=guide_legend(keywidth = 3, keyheight = 1,nrow=2,byrow=TRUE))

```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



```

ggsave( paste0("figures/", "fig16band17a.png"), width = 16, height = 9, bg = "white",
scale = .4)

```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

```
rm(dt_p); rm(dt_citing_cited_journals); rm(dt_plot); rm(dt_plot_angrist)
```

For the third panel of figure 1 we did things differently. Since we were not able to identify, in our database, every journals Angrist and al. were mapping in their ‘Other Disciplines’ category, we took every journals that were in those disciplines in our set of category. Angrist’s ‘Other Disciplines’ are Computer science, Mathematics, Statistics, Medicine, Operations, Research, Physics and Public Health.

For our first attempt, we took every medicine specialities. This choice led the proportion of COC from psychology to ‘Other Disciplines’ to go up to almost 20%. This is because psychology cites Neurology & Neurosurgery and Psychiatry a lot. In Angrist et al’s list of medicine journals, there are one psychiatry journal, and two neurology journals out of approximately 75 journals. As we can see in the following code, if we look at what the top 5% psychology journals cites, we find that it cites 422 different Psychiatry or Neurology & Neurosurgery journals (according to our categories). Since we don’t know which of those 428

journals are the ones in Angrist's list, we decided not to track Psychiatry and Neurology altogether. With this restriction, we get almost the same results as Angrist et al.

```
dt_citing_cited_journals <- readRDS(paste0(path_to_project_data,
                                           "dt_citing_cited_journals.rds" ))
load(paste0(path_to_project_data, "weight_angrist.RData"))
setnames(dt_citing_cited_journals, "Year", "citing_Year")
dt_citing_cited_journals <- merge(dt_citing_cited_journals,
                                 weight_angrist,
                                 by.x = c("Discipline_citing", "citing_code_revue",
                                           "citing_Year"),
                                 by.y = c("Discipline_citing", "cited_code_revue",
                                           "citing_Year"))

dt_psy <- dt_citing_cited_journals[Discipline_citing == "Psychology" &
                                   between(citing_Year, 1970, 2015)]

dt_psy <- data.table(group_by(dt_psy, citing_Year) %>%
                    top_frac(n=0.05, wt=discipline_proportion))
uniqueN(dt_psy[cited_discipline %in% c(51,60)]$cited_code_revue)

## [1] 422
rm(dt_citing_cited_journals, dt_psy)
```

In the last step, we keep Angrist et al's weights, but we drop the restriction regarding the disciplines considered as recipients of citations:

```
dt_citing_cited_journals <- readRDS(paste0(path_to_project_data,
                                           "dt_citing_cited_journals.rds" ))
load(paste0(path_to_project_data, "weight_angrist.RData"))
setnames(dt_citing_cited_journals, "Year", "citing_Year")
#Renaming discipline to match Angrist's category names
dt_citing_cited_journals <- renaming_ssh_angrist(dt_citing_cited_journals)

#Taking relevant columns and summing number of citation going
# from every discipline to every other (including itself)
dt_p <- dt_citing_cited_journals[!is.na(Discipline_cited) &
                                 Discipline_citing %in% core_angrist,
                                 list(citing_code_revue, Discipline_citing,
                                       Discipline_cited, citing_Year, N)][
                                 ,sum(N), by = list(citing_code_revue,
                                                    Discipline_citing,
                                                    Discipline_cited,
                                                    citing_Year) ]

setnames(dt_p, "V1", "Nb_cited")

#Merging Angrist's weights
dt_p <- merge(dt_p, weight_angrist, by.x = c("Discipline_citing",
                                           "citing_code_revue", "citing_Year"),
             by.y = c("Discipline_citing", "cited_code_revue", "citing_Year"))

setnames(dt_p, "discipline_proportion", "Journal_weight_in_disc")
## Calculating proportions of extradisciplinary citations
dt_p[, Total_journal := sum(Nb_cited), by = .(citing_Year, citing_code_revue)][
  , Proportion := Nb_cited / Total_journal][
```



```

, := `(Nb_cited = NULL, Total_journal = NULL)]
## Calculating weighted proportions
dt_p[, weighted_proportion := Journal_weight_in_disc * Proportion] [
, := `(Journal_weight_in_disc = NULL, Proportion = NULL)]

# Getting discipline names for Social Sciences graph and Other Disciplin graph
SSH_without_management <- unique(dt_citing_cited_journals[
  citing_discipline > 100 & Discipline_citing != 132]$Discipline_citing)
NSE <- unique(dt_citing_cited_journals[citing_discipline <= 100]$Discipline_citing)

#This chunk of code make it possible to have three graphs next to each other in ggplot
dt_p[Discipline_cited %in% SSH_without_management, Graph := "1. Social Sciences"]
dt_p[Discipline_cited == "Management", Graph := "2. Business Disciplines"]
dt_p[Discipline_cited %in% NSE, Graph := "3. Other Disciplines"]
dt_p <- dt_p[!is.na(Graph)]

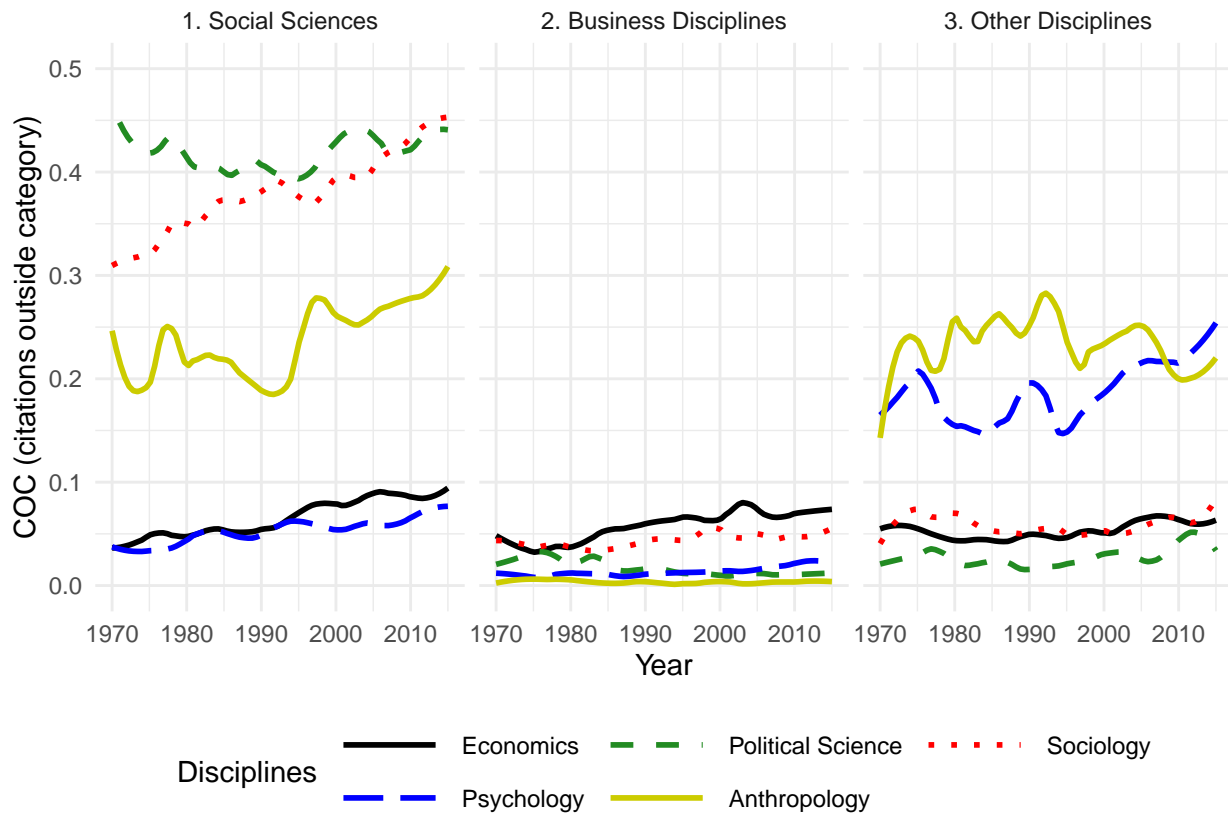
## Summing weighted proportions
dt_plot_angrist <- dt_p[Discipline_citing != Discipline_cited, sum(weighted_proportion),
  by = .(Discipline_citing, Graph, citing_Year)]
setnames(dt_plot_angrist, "V1", "Proportion")
order_disc <- c("Economics", "Political Science", "Sociology", "Psychology", "Anthropology")
dt_plot_angrist$Discipline_citing <- factor(dt_plot_angrist$Discipline_citing,
  levels = order_disc)
dt_plot_angrist <- dt_plot_angrist[between(citing_Year, 1970, 2015)]

#colors <- c("Psychology" = "blue", "Economics" = "black", "Sociology" = "red",
#           "Anthropology" = "#CCCC00", "Political Science" = "#228B22")
#linetypes <- c("Psychology" = "longdash", "Economics" = "solid",
#              "Sociology" = "dotted", "Anthropology" = "solid",
#              "Political Science" = "dashed")

ggplot(dt_plot_angrist, aes(x=citing_Year, y=Proportion, color=Discipline_citing)) +
  geom_smooth(se = F, span = 0.25, aes(linetype = Discipline_citing)) +
  labs(x="Year", y="COC (citations outside category)", col="Discipline") +
  ylim(0, 0.5) +
  # ggtitle("Fig. 1 reproduction with exhaustive categories") +
  facet_wrap(~ Graph, scales = "free_x") + theme_minimal() +
  theme(legend.position="bottom") +
  scale_color_manual(name = "Disciplines", values = colors_angrist) +
  scale_linetype_manual(name = "Disciplines", values = linetypes_angrist) +
  guides(linetype=guide_legend(keywidth = 3, keyheight = 1, nrow=2, byrow=TRUE),
         color=guide_legend(keywidth = 3, keyheight = 1, nrow=2, byrow=TRUE))

## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).

```



```
ggsave( paste0("figures/", "fig17b.png"), width = 16, height = 9, bg = "white",
         scale = .4)
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

```
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```

```
rm(dt_p, dt_citing_cited_journals, dt_plot_angrist)
```

The change is radical: in this new version, economics is the only discipline that remains below the 10% line in all panels.