

# STM Topicmodel

Riccardo Ruta

05/2022

## Contents

<b>STM Topic model analysis</b>	<b>1</b>
Preliminary steps . . . . .	1
FIND THE BEST NUMBER OF TOPICS K . . . . .	3
Coefficients . . . . .	11
Interpretation and validation . . . . .	17

## STM Topic model analysis

### Preliminary steps

#### Load the data

#### Import the dictionaries

```
# Import dictionaries file
dict <- read_excel("data/populism_dictionaries.xlsx")

Decadri_Boussalis_Grundl <-
  dictionary(list(people =
    dict$Decadri_Boussalis_Grundl_People
    [!is.na(dict$Decadri_Boussalis_Grundl_People)],
    common_will =
    dict$`Decadri_Boussalis_Grundl_Common Will`
    [!is.na(dict$`Decadri_Boussalis_Grundl_Common Will`)],
    elite =
    dict$Decadri_Boussalis_Grundl_Elite
    [!is.na(dict$Decadri_Boussalis_Grundl_Elite)]))
```

#### Remove all the account's mentions

```
DFM@Dimnames$features <- gsub("^@", "", DFM@Dimnames$features)
```

## Trim the data

```
# Remove text with less than 1 word
DFM <- dfm_subset(DFM, ntoken(DFM) > 1)

# Remove very short words
DFM <- dfm_remove(DFM, min_nchar=2)

# Dfm trimming: only words that occur in the top 20% of the distribution and in less than 10% of documents
DFM <- dfm_trim(DFM, min_termfreq = 0.8,
               termfreq_type = "quantile",
               max_docfreq = 0.1, docfreq_type = "prop")
```

## Group and weight the data

```
DFMG<-dfm_group(DFM, groups = interaction(nome, quarter, party_id))

DFMGW<- dfm_weight(DFMG,
                  scheme ="prop")
```

## Apply dictionary

```
# Apply Dictionary
DFMdict <- dfm_lookup(DFMGW, dictionary = Decadri_Boussalis_Grundl)

# Convert to a dataframe
DATAdictDFM <- DFMdict %>%
  quantda::convert(to = "data.frame")
```

## Create percentage for each components

```
# RUN ONLY ONCE!
# Add variable with general level of populism & multiply all components by 100
DATAdictDFM <- DATAdictDFM %>% mutate(populism = (people + common_will + elite) * 100)

DATAdictDFM <- DATAdictDFM %>% mutate(people = people*100,
                                       common_will = common_will*100,
                                       elite = elite*100)
```

## Add the percentage of populism to the original dfm (not weighted)

```
docvars(DFMG) <- cbind(docvars(DFMG), DATAdictDFM)
```

## Convert DFM to STM format

```
#Extract a sample
myDFM <- dfm_sample(DFMG,
#                               size = 5000)

myDFM = DFMG
set.seed(123)
DfmStm <- quanteda::convert(myDFM, to = "stm", docvars = docvars(myDFM))
```

```
## Warning in dfm2stm(x, docvars, omit_empty = TRUE): Dropped empty document(s): DE
## LORENZIS Diego.7.M5S, DI MICCO Fabio.8.M5S, MARZANA Maria.9.M5S
```

## FIND THE BEST NUMBER OF TOPICS K

Search the best number of Topics comparing coherence and exclusivity values

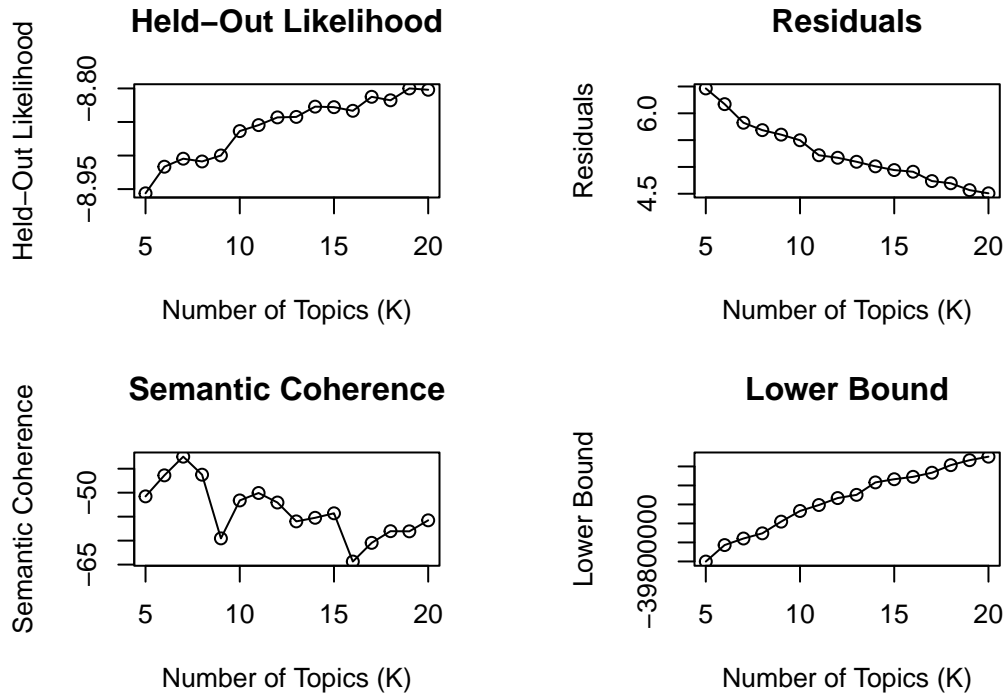
K = 2:50

```
k <-c(5:20)
system.time(storage <- searchK(DfmStm$documents,
                              DfmStm$vocab,
                              K = k,
                              #max.em.its = 75, # REMOVED
                              prevalence = ~ party_id + populism + s(quarter),
                              data = DfmStm$meta, init.type = "Spectral"))
#save(storage, file="data/storage.Rda")
```

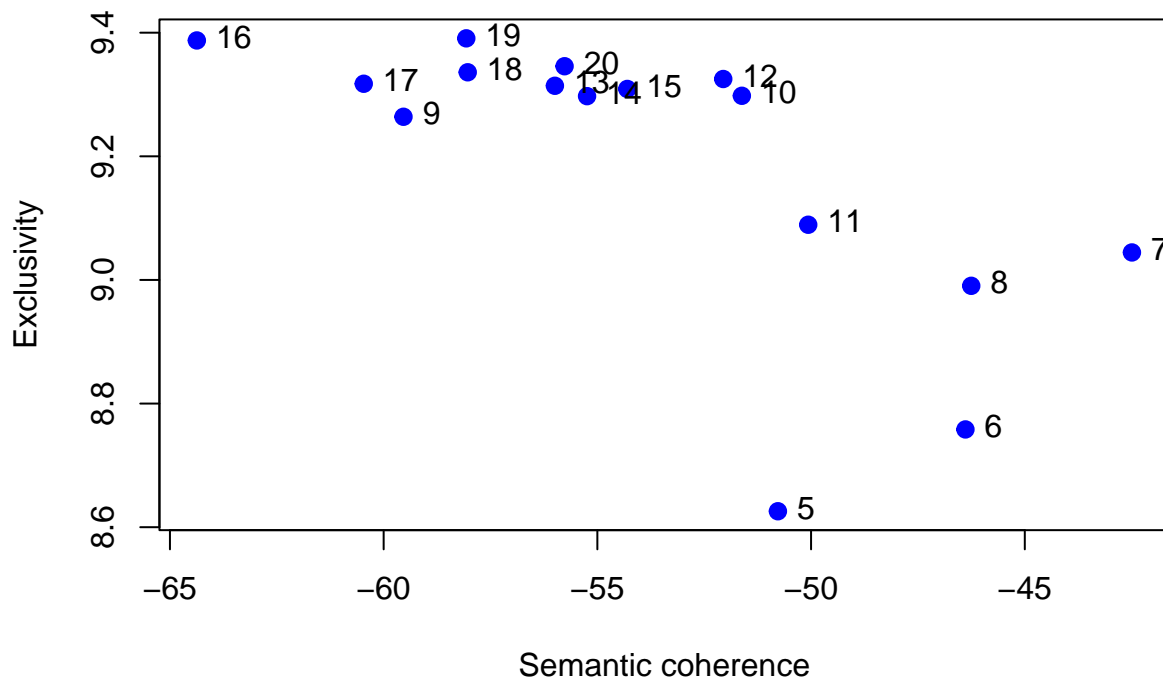
plot results

```
plot.searchK(storage)
```

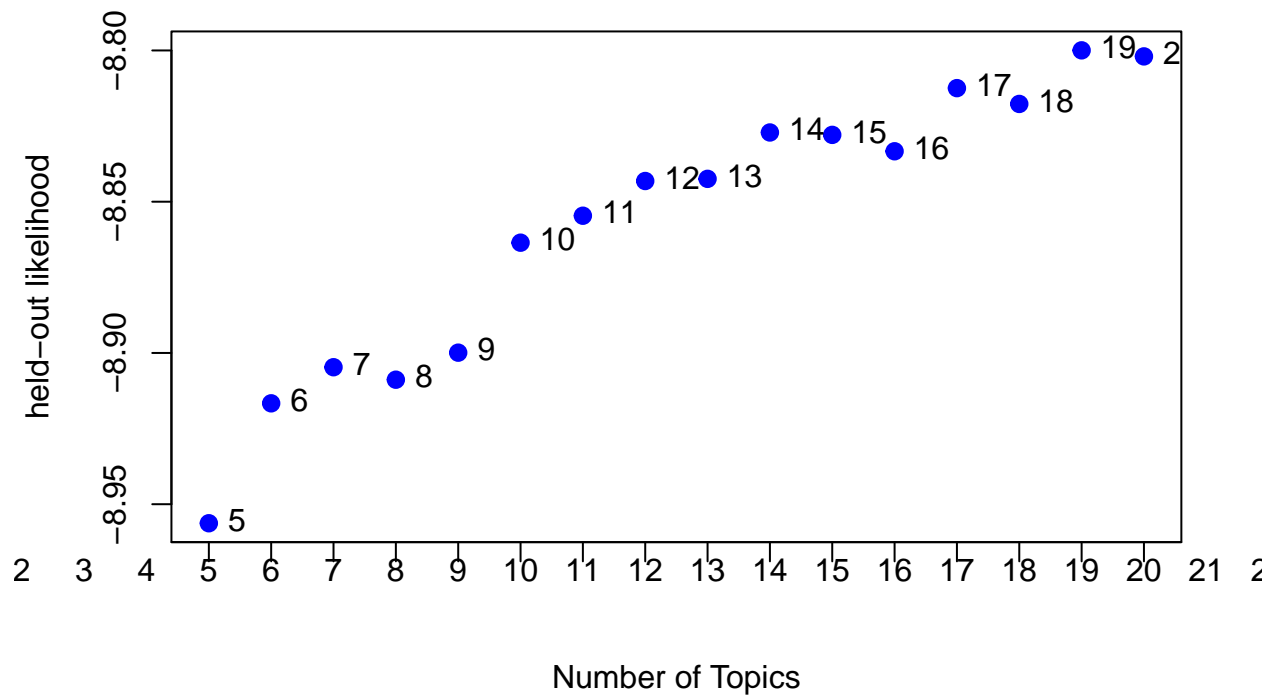
## Diagnostic Values by Number of Topics



```
plot(storage$results$semcoh, storage$results$exclus,
      xlab= "Semantic coherence",
      ylab= "Exclusivity",
      col= "blue", pch = 19, cex = 1, lty = "solid", lwd = 2)
text(storage$results$semcoh, storage$results$exclus, labels=storage$results$K, cex= 1, pos=4)
```



```
plot(storage$results$K, storage$results$heldout,
      xlab= "Number of Topics",
      ylab= "held-out likelihood",
      col= "blue", pch = 19, cex = 1, lty = "solid", lwd = 2, xaxt="n")
text(storage$results$K, storage$results$heldout, labels=storage$results$K, cex= 1, pos=4)
xtick<-seq(2, 50, by=1)
axis(side=1, at=xtick, labels = FALSE)
text(x=xtick, par("usr")[3],
      labels = xtick, pos = 1, xpd = TRUE)
```



K= 11 has the best values of coherence, exclusivity and held-Out likelihood.

Run the analysis selecting k = 11

```
k = 11

mySTM <- stm(DfmStm$documents, vocab = DfmStm$vocab,
             K = k,
             prevalence = ~ party_id + populism + s(quarter),
             data = DfmStm$meta,
             init.type = "Spectral",
             verbose = TRUE)

# save(mySTM, file="data/mySTM.Rda")
```

## Label topics

The frequency/exclusivity (FREX) scoring summarizes words according to their probability of appearance under a topic and the exclusivity to that topic. These words provide more semantically intuitive representations of each topic

```
labelTopics(mySTM, n=10)
```

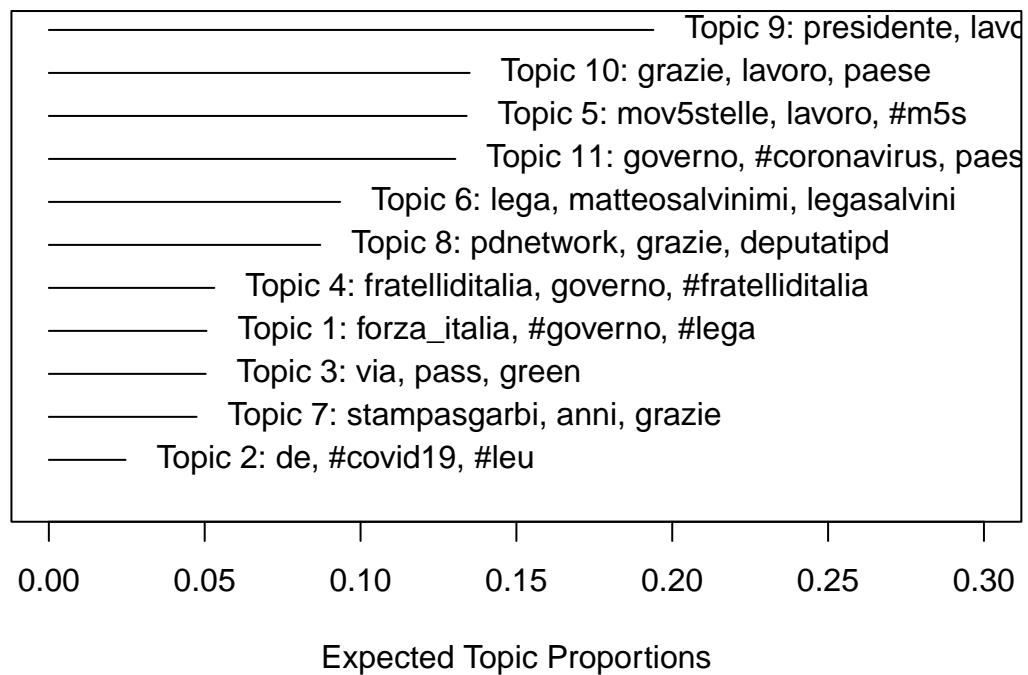
## Topic 1 Top Words:  
 ## Highest Prob: forza\_italia, #governo, #lega, gruppoficamera, via, grazie, presidente, #italia,  
 ## FREX: gruppoficamera, patriziarametta, pittoni, gruppofisenato, votalega, #genteevalori, #lega  
 ## Lift: #grimoldi, normagi2, votalega, #abccidanielasbrollini, #alternativo, #barbagia, #bergama  
 ## Score: forza\_italia, gruppoficamera, pittoni, patriziarametta, #legasardegna, a\_gusmeroli, vota  
 ## Topic 2 Top Words:  
 ## Highest Prob: de, #covid19, #leu, italymfa, en, l'aggiornamento, bollettino, el, crisi, et  
 ## FREX: rapite, dall'oglio, #padredalloglio, paoladelusa, cooperazione\_it, #sahel, #italianinelm  
 ## Lift: #2030isnow, #africaeurope, #bamako, #ciudadania, #cnsc, #corsaacasa, #corsamattutina, #c  
 ## Score: #padredalloglio, rapite, paoladelusa, cooperazione\_it, dall'oglio, #leu, #maie, en, l'a  
 ## Topic 3 Top Words:  
 ## Highest Prob: via, pass, green, draghi, rai, governo, fattoquotidiano, anni, vaccinati, conte  
 ## FREX: gfi65, adginforma, massionline, angelazoppo, dottorbarbieri, andreagiuricin, #noobbligov  
 ## Lift: #datigrezzi, #laneuro, #manif28luglio, #prelemi, #rey, alternativa\_it, ballifrancesco, b  
 ## Score: anzaldi, gfi65, massionline, adginforma, angelazoppo, ugambini, gianni\_dragoni, red\_mar  
 ## Topic 4 Top Words:  
 ## Highest Prob: fratelliditalia, governo, #fratelliditalia, italiani, via, giorgiameloni, meloni  
 ## FREX: #fratelliditalia, vocedelpatriota, #giorgiameloni, #fdi, fratelliditaia, fratelliditali  
 ## Lift: #orgogliotricolore, #patrioti, -staff, delmastro, #accalarentia, #aisha, #arcuridimettit  
 ## Score: fratelliditalia, vocedelpatriota, #fratelliditalia, fdi\_parlamento, #giorgiameloni, fra  
 ## Topic 5 Top Words:  
 ## Highest Prob: mov5stelle, lavoro, #m5s, paese, legge, diretta, cittadini, commissione, grazie,  
 ## FREX: m5s\_senato, mov5stelle, #superbonus, #transizioneecologica, a\_lisacorrado, m5s\_camera, m  
 ## Lift: #acqua2050, #annamariaparente, #anpr, #anticorruzione, #beigua, #biellese, #canapa, #cis  
 ## Score: mov5stelle, m5s\_senato, #m5s, a\_lisacorrado, greenitalia1, puglia\_m5s, m5scampania, fra  
 ## Topic 6 Top Words:  
 ## Highest Prob: lega, matteosalvinimi, legasalvini, #lega, #salvini, salvini, grazie, governo, m  
 ## FREX: lega\_senato, angelociocca, #processateancheme, #oggiavotolega, #primalitalia, #26gennaioov  
 ## Lift: #accademiafederalelega, #aiutiamociacasanostra, #alballottaggiovotolega, #autonomiasubito  
 ## Score: #iostoconsalvini, legacamera, lega\_senato, legasalvini, matteosalvinimi, angelociocca, s  
 ## Topic 7 Top Words:  
 ## Highest Prob: stampasgarbi, anni, grazie, città, regione, grande, #veneziana, veneto, #liguria, a  
 ## FREX: stampasgarbi, comuneveneziana, #governomusumeci, nino\_ippolito, regione\_sicilia, cmveneziana  
 ## Lift: #venetodaamare, #abruzzesi, #antimafiafuffa, #baretta sindaco, #buonasalute, #campalto, #  
 ## Score: stampasgarbi, #governomusumeci, nino\_ippolito, comuneveneziana, #orgoglioveneto, #venetoda  
 ## Topic 8 Top Words:  
 ## Highest Prob: pdnetwork, grazie, deputatipd, politica, bene, anni, legge, governo, pd, forza\_i  
 ## FREX: azione\_it, nomfup, piu\_europa, graziano\_delrio, elevisconti, adalucde, curini, giusvapule  
 ## Lift: #cocaweb, #colapescedimartino, #dopofestival, #nanniespresso, #neverstop, #pensieridemoc  
 ## Score: deputatipd, azione\_it, pdnetwork, senatoripd, nomfup, giusvapulejo, di\_reddito, piu\_eur  
 ## Topic 9 Top Words:  
 ## Highest Prob: presidente, lavoro, paese, grazie, grande, anni, donne, l'italia, politica, solia  
 ## FREX: #ucraina, #patrickzaki, #romarinasce, #putin, legadilettanti, ucraino, #italiateam, #gua  
 ## Lift: #100m, #25novembre2021, #afirenzeperlapace, #agorademocratiche, #altracosa, #angelamerke  
 ## Score: italiaviva, #romarinasce, pdnetwork, legadilettanti, #tokyo2020, #gualtierisindaco, #pa  
 ## Topic 10 Top Words:  
 ## Highest Prob: grazie, lavoro, paese, grande, cittadini, anni, #covid19, momento, presidente, s  
 ## FREX: gdf, #calabria, \_carabinieri\_, #palermo, esprimo, poliziadistato, #gianlucarospi, cordog  
 ## Lift: #massimotroisi, #materata2019, edraspa, tonello, #11giugno, #25giugno, #ioestolibero, #pop  
 ## Score: #covid19, #gianlucarospi, #popoloprotagonista, \_carabinieri\_, #calabria, gt, #sud, #riap  
 ## Topic 11 Top Words:  
 ## Highest Prob: governo, #coronavirus, paese, italiani, italia, l'italia, politica, imprese, con  
 ## FREX: #mes, mes, liquidità, opposizioni, #prescrizione, #italiaviva, #coronavirus, #covid2019,  
 ## Lift: #29febbraio, #annibali, #diamondprincess, #gennaroarma, #lamossadelcavallo, #pianoxitalia

```
##      Score: #coronavirus, italiaviva, #mes, mes, #fase2, liquidità, #covid2019, imprese, #conte, #p
```

Most frequent topic

```
R <- plot(mySTM,  
          type = "summary",  
          xlim = c(0, .3))
```

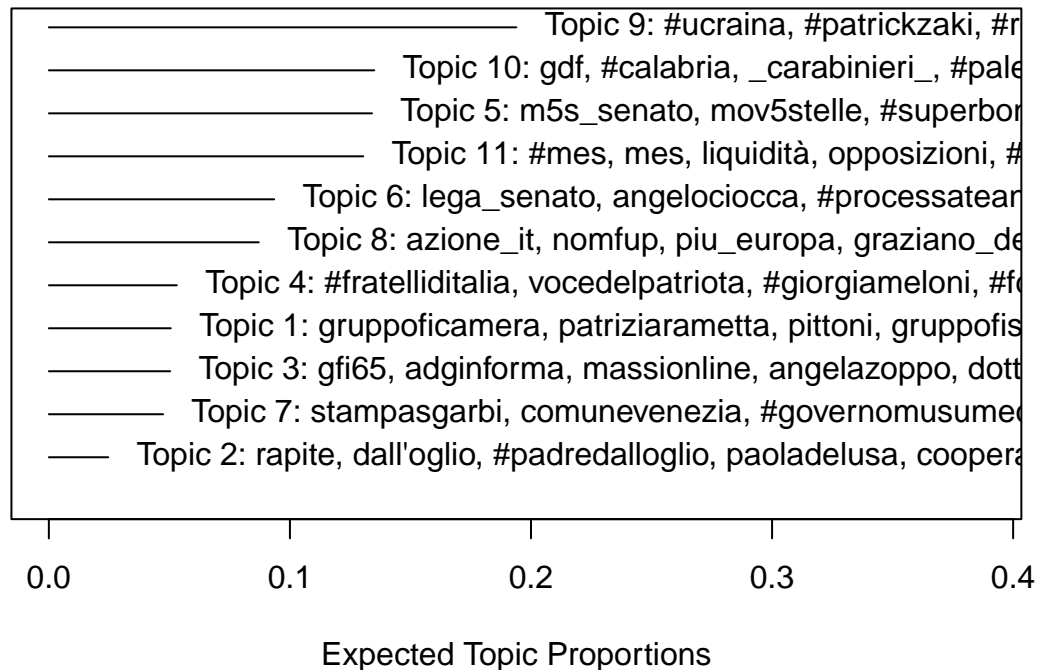
## Top Topics



```
# plot just frex words for each topic  
plot(mySTM, type = "summary", labeltype = c("frex"), n=5) # topic 9 is the most frequent
```



## Top Topics



Find document most associated Text with the most frequent topic (9)

```
# Load original corpus
load("data/corpus.Rda")

# list the documents in the dfm
docs <- myDFM@Dimnames$docs

# Remove text with less than 1 word
corpus <- corpus_subset(corpus, ntoken(corpus) > 1)

# group the corpus like the dfm
corpus_g <- corpus_group(corpus, groups = interaction(nome, quarter, party_id))

# subset the same text of the dfm
subs_corpus <- corpus_subset(corpus_g, docnames(corpus_g) %in% docs)

documents <- as.character(subs_corpus)

documents <- as.vector(documents)

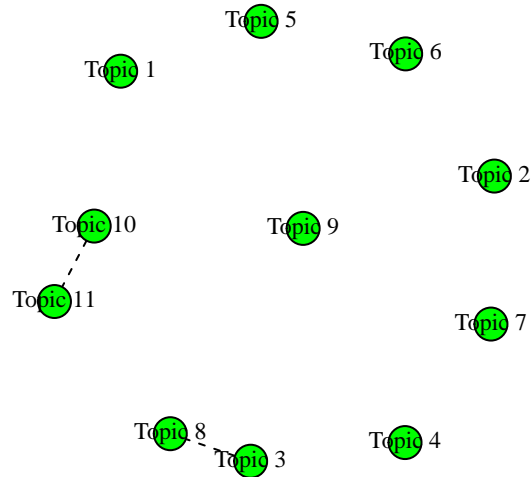
# Let's focus on topic 9
thought9 <- findThoughts(mySTM, texts=documents, topics=9, n=3)$docs[[1]]
```

```
# qui non run esce errore seguente !!

# Error in findThoughts(mySTM, texts = texts(subs_corpus), topics = 9, n = 3) :
# Number of provided texts and number of documents modeled do not match
```

## Correlation between topics

```
mod.out.corr <- topicCorr(mySTM)
plot(mod.out.corr)
```



Which are the the most likely topics across our documents?

```
#apply(mySTM$theta, 1, which.max)

tab <- table(apply(mySTM$theta, 1, which.max))
kable(tab[order(desc(tab))])
```

Var1	Freq
9	1309
5	809
11	707
10	652
6	631
8	463
4	309
1	245
3	245
7	207
2	133

save them back in the original dataframe

```
# STESSO PROBLEMA LIKE SOPRA:
# ORIGINAL CORPUS E STM NON SONO LO STESSO NUMERO
# 5710 vs 5713
subs_corpus$topic <- apply(mySTM$theta,1,which.max)

str(subs_corpus)

# Topic 5 - 5 random documents associated to it

set.seed(123)

sample(subs_corpus$text[subs_corpus$topic==5], 5)
```

## Coefficients

```
#out$meta$rating <- as.factor(out$meta$rating)
prep <- estimateEffect(1:11 ~ party_id + populism + s(quarter),
                      mySTM,metadata = DfmStm$meta,
                      uncertainty = "Global")
summary(prep)
```

```
##
## Call:
## estimateEffect(formula = 1:11 ~ party_id + populism + s(quarter),
##               stmobj = mySTM, metadata = DfmStm$meta, uncertainty = "Global")
##
##
## Topic 1:
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.190725   0.018447  10.339 < 2e-16 ***
## party_idFDI   -0.140034   0.020417  -6.859 7.68e-12 ***
## party_idFI    -0.011609   0.018405  -0.631  0.5282
```

```

## party_idINDIPENDENTE -0.171896 0.028309 -6.072 1.34e-09 ***
## party_idIV -0.124675 0.030036 -4.151 3.36e-05 ***
## party_idLEGA -0.085065 0.018706 -4.547 5.54e-06 ***
## party_idLEU -0.148444 0.022095 -6.719 2.01e-11 ***
## party_idM5S -0.141948 0.018119 -7.834 5.59e-15 ***
## party_idMISTO -0.111554 0.018983 -5.877 4.43e-09 ***
## party_idPD -0.154919 0.017724 -8.741 < 2e-16 ***
## party_idREG_LEAGUES -0.152064 0.028152 -5.402 6.88e-08 ***
## populism -0.008643 0.001523 -5.677 1.44e-08 ***
## s(quarter)1 0.051605 0.132457 0.390 0.6968
## s(quarter)2 0.004390 0.058056 0.076 0.9397
## s(quarter)3 -0.003873 0.030220 -0.128 0.8980
## s(quarter)4 -0.017871 0.019907 -0.898 0.3694
## s(quarter)5 -0.029209 0.014478 -2.017 0.0437 *
## s(quarter)6 -0.021903 0.019411 -1.128 0.2592
## s(quarter)7 -0.018924 0.029885 -0.633 0.5266
## s(quarter)8 -0.036529 0.034163 -1.069 0.2850
## s(quarter)9 -0.018668 0.010689 -1.747 0.0808 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 2:
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0240963 0.0137979 1.746 0.080800 .
## party_idFDI -0.0038512 0.0146415 -0.263 0.792536
## party_idFI -0.0028595 0.0136101 -0.210 0.833599
## party_idINDIPENDENTE 0.1336166 0.0299178 4.466 8.12e-06 ***
## party_idIV 0.0039485 0.0247786 0.159 0.873399
## party_idLEGA -0.0054179 0.0132503 -0.409 0.682639
## party_idLEU -0.0072656 0.0178174 -0.408 0.683449
## party_idM5S 0.0154623 0.0129847 1.191 0.233778
## party_idMISTO 0.0533434 0.0147772 3.610 0.000309 ***
## party_idPD 0.0226076 0.0131485 1.719 0.085597 .
## party_idREG_LEAGUES 0.1451007 0.0355056 4.087 4.44e-05 ***
## populism -0.0068598 0.0014474 -4.739 2.20e-06 ***
## s(quarter)1 0.0474909 0.1095867 0.433 0.664767
## s(quarter)2 -0.0254836 0.0479738 -0.531 0.595302
## s(quarter)3 0.0096613 0.0265692 0.364 0.716149
## s(quarter)4 -0.0010323 0.0181295 -0.057 0.954596
## s(quarter)5 -0.0008385 0.0120987 -0.069 0.944748
## s(quarter)6 0.0015696 0.0183169 0.086 0.931716
## s(quarter)7 -0.0065401 0.0279764 -0.234 0.815170
## s(quarter)8 0.0077881 0.0335693 0.232 0.816546
## s(quarter)9 0.0063163 0.0102118 0.619 0.536254
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 3:
##
## Coefficients:

```

```

##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.0092435  0.0156795   0.590  0.5555
## party_idFDI      -0.0060580  0.0171258  -0.354  0.7236
## party_idFI       0.0102323  0.0155696   0.657  0.5111
## party_idINDIPENDENTE -0.0498332  0.0276797  -1.800  0.0719 .
## party_idIV       -0.0213460  0.0271226  -0.787  0.4313
## party_idLEGA      0.0205457  0.0160654   1.279  0.2010
## party_idLEU      -0.0231779  0.0203157  -1.141  0.2540
## party_idM5S       0.0132186  0.0149348   0.885  0.3761
## party_idMISTO     0.0754295  0.0157202   4.798 1.64e-06 ***
## party_idPD       -0.0101944  0.0152452  -0.669  0.5037
## party_idREG_LEAGUES -0.0328108  0.0265988  -1.234  0.2174
## populism         0.0009841  0.0017731   0.555  0.5789
## s(quarter)1      -0.0203254  0.1291961  -0.157  0.8750
## s(quarter)2       0.0149011  0.0569990   0.261  0.7938
## s(quarter)3      -0.0105891  0.0311940  -0.339  0.7343
## s(quarter)4       0.0303174  0.0206005   1.472  0.1412
## s(quarter)5       0.0119613  0.0141149   0.847  0.3968
## s(quarter)6       0.1445508  0.0251046   5.758 8.96e-09 ***
## s(quarter)7       0.0446011  0.0358703   1.243  0.2138
## s(quarter)8       0.1047823  0.0431660   2.427  0.0152 *
## s(quarter)9       0.0625024  0.0118577   5.271 1.41e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 4:
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.0178879  0.0111110  -1.610 0.107471
## party_idFDI      0.4513405  0.0174825  25.817 < 2e-16 ***
## party_idFI       0.0454711  0.0113961   3.990 6.69e-05 ***
## party_idINDIPENDENTE -0.0062589  0.0208235  -0.301 0.763755
## party_idIV       -0.0063568  0.0193566  -0.328 0.742619
## party_idLEGA     0.0170733  0.0110033   1.552 0.120801
## party_idLEU      0.0169227  0.0152916   1.107 0.268485
## party_idM5S      0.0107626  0.0104781   1.027 0.304394
## party_idMISTO     0.0415945  0.0117438   3.542 0.000401 ***
## party_idPD       0.0029460  0.0105926   0.278 0.780934
## party_idREG_LEAGUES 0.0098415  0.0207680   0.474 0.635605
## populism        0.0156521  0.0021551   7.263 4.31e-13 ***
## s(quarter)1      0.0323322  0.0995359   0.325 0.745322
## s(quarter)2     -0.0031862  0.0438591  -0.073 0.942090
## s(quarter)3      0.0321969  0.0248566   1.295 0.195268
## s(quarter)4      0.0641008  0.0165334   3.877 0.000107 ***
## s(quarter)5      0.0257806  0.0108376   2.379 0.017401 *
## s(quarter)6     -0.0073199  0.0161843  -0.452 0.651083
## s(quarter)7      0.0659624  0.0235743   2.798 0.005158 **
## s(quarter)8     -0.0633661  0.0277408  -2.284 0.022395 *
## s(quarter)9      0.0005911  0.0081236   0.073 0.942001
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

```

##
## Topic 5:
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.034194   0.018880   1.811 0.070186 .
## party_idFDI    -0.003332   0.019455  -0.171 0.864036
## party_idFI      0.023307   0.017842   1.306 0.191498
## party_idINDIPENDENTE 0.215220   0.042256   5.093 3.63e-07 ***
## party_idIV      0.098809   0.034980   2.825 0.004748 **
## party_idLEGA    0.003717   0.017715   0.210 0.833800
## party_idLEU     0.137799   0.026714   5.158 2.58e-07 ***
## party_idM5S     0.235287   0.017411  13.514 < 2e-16 ***
## party_idMISTO   0.071424   0.018370   3.888 0.000102 ***
## party_idPD      0.047387   0.017280   2.742 0.006119 **
## party_idREG_LEAGUES 0.042825   0.037211   1.151 0.249825
## populism       -0.007342   0.002234  -3.286 0.001022 **
## s(quarter)1    -0.087316   0.158631  -0.550 0.582044
## s(quarter)2     0.013665   0.070109   0.195 0.845473
## s(quarter)3     0.060733   0.037687   1.611 0.107130
## s(quarter)4    -0.055346   0.023022  -2.404 0.016246 *
## s(quarter)5     0.053370   0.020477   2.606 0.009177 **
## s(quarter)6    -0.058525   0.026149  -2.238 0.025252 *
## s(quarter)7     0.209564   0.042837   4.892 1.02e-06 ***
## s(quarter)8    -0.192113   0.047584  -4.037 5.48e-05 ***
## s(quarter)9     0.007735   0.014751   0.524 0.600056
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 6:
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.042102   0.015707   2.681 0.007372 **
## party_idFDI    -0.016551   0.016486  -1.004 0.315470
## party_idFI     -0.011165   0.015085  -0.740 0.459231
## party_idINDIPENDENTE -0.025420   0.028086  -0.905 0.365473
## party_idIV     -0.021646   0.025832  -0.838 0.402075
## party_idLEGA    0.385872   0.016192  23.830 < 2e-16 ***
## party_idLEU    -0.016974   0.019687  -0.862 0.388606
## party_idM5S     0.001058   0.014302   0.074 0.941015
## party_idMISTO  -0.012542   0.015373  -0.816 0.414614
## party_idPD     -0.015042   0.014533  -1.035 0.300700
## party_idREG_LEAGUES -0.001621   0.027411  -0.059 0.952857
## populism       0.012548   0.002292   5.475 4.57e-08 ***
## s(quarter)1    -0.082477   0.142831  -0.577 0.563663
## s(quarter)2    -0.010643   0.061357  -0.173 0.862300
## s(quarter)3     0.040077   0.035015   1.145 0.252442
## s(quarter)4    -0.075153   0.020712  -3.628 0.000288 ***
## s(quarter)5     0.030592   0.017330   1.765 0.077567 .
## s(quarter)6    -0.034046   0.022979  -1.482 0.138488
## s(quarter)7    -0.052179   0.035092  -1.487 0.137093
## s(quarter)8    -0.058081   0.037991  -1.529 0.126370

```

```

## s(quarter)9          -0.040076   0.011951  -3.353 0.000804 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 7:
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.569e-01  1.996e-02   7.860 4.56e-15 ***
## party_idFDI    -7.637e-02  2.233e-02  -3.420 0.000631 ***
## party_idFI     -8.308e-02  2.070e-02  -4.013 6.08e-05 ***
## party_idINDIPENDENTE 1.114e-01  3.687e-02   3.021 0.002528 **
## party_idIV     -1.345e-01  2.925e-02  -4.598 4.36e-06 ***
## party_idLEGA   -7.708e-02  2.011e-02  -3.832 0.000128 ***
## party_idLEU    -1.193e-01  2.445e-02  -4.879 1.10e-06 ***
## party_idM5S    -1.209e-01  1.971e-02  -6.135 9.11e-10 ***
## party_idMISTO   -1.070e-01  2.026e-02  -5.279 1.35e-07 ***
## party_idPD     -9.431e-02  1.975e-02  -4.775 1.84e-06 ***
## party_idREG_LEAGUES 1.899e-02  3.968e-02   0.479 0.632293
## populism       -8.086e-03  1.619e-03  -4.993 6.12e-07 ***
## s(quarter)1    -2.218e-01  1.252e-01  -1.771 0.076594 .
## s(quarter)2     5.329e-02  5.504e-02   0.968 0.333044
## s(quarter)3     1.650e-03  3.030e-02   0.054 0.956576
## s(quarter)4    -3.088e-02  1.969e-02  -1.569 0.116801
## s(quarter)5    -1.975e-05  1.590e-02  -0.001 0.999009
## s(quarter)6     6.910e-02  2.193e-02   3.151 0.001636 **
## s(quarter)7    -6.219e-02  3.236e-02  -1.922 0.054712 .
## s(quarter)8     1.404e-02  3.633e-02   0.387 0.699073
## s(quarter)9     2.731e-02  1.177e-02   2.320 0.020360 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 8:
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.084574  0.019215   4.401 1.1e-05 ***
## party_idFDI    -0.005658  0.020531  -0.276 0.782861
## party_idFI     -0.014923  0.019153  -0.779 0.435936
## party_idINDIPENDENTE -0.098924  0.033176  -2.982 0.002878 **
## party_idIV     -0.012537  0.034729  -0.361 0.718115
## party_idLEGA   -0.059670  0.019214  -3.106 0.001908 **
## party_idLEU     0.057184  0.027071   2.112 0.034697 *
## party_idM5S    -0.050148  0.018381  -2.728 0.006385 **
## party_idMISTO   0.022403  0.020011   1.120 0.262966
## party_idPD     0.067000  0.019400   3.454 0.000557 ***
## party_idREG_LEAGUES -0.034375  0.034465  -0.997 0.318624
## populism       0.001519  0.001951   0.779 0.436061
## s(quarter)1    -0.373088  0.152163  -2.452 0.014241 *
## s(quarter)2     0.169888  0.066832   2.542 0.011047 *
## s(quarter)3    -0.072317  0.035999  -2.009 0.044602 *
## s(quarter)4     0.063879  0.024315   2.627 0.008633 **

```

```

## s(quarter)5          0.007634    0.017982    0.425 0.671195
## s(quarter)6          0.008651    0.025218    0.343 0.731566
## s(quarter)7          0.029230    0.040972    0.713 0.475621
## s(quarter)8          0.016353    0.046882    0.349 0.727246
## s(quarter)9          0.003070    0.013391    0.229 0.818663
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 9:
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.079964   0.022235   3.596 0.000325 ***
## party_idFDI    -0.120279   0.023495  -5.119 3.17e-07 ***
## party_idFI     -0.022564   0.022083  -1.022 0.306925
## party_idINDIPENDENTE -0.040564   0.048447  -0.837 0.402465
## party_idIV      0.140044   0.044230   3.166 0.001552 **
## party_idLEGA   -0.112169   0.021543  -5.207 1.99e-07 ***
## party_idLEU     0.074285   0.030973   2.398 0.016499 *
## party_idM5S    -0.024584   0.021314  -1.153 0.248786
## party_idMISTO   -0.039088   0.023802  -1.642 0.100595
## party_idPD      0.133881   0.023085   5.800 7.01e-09 ***
## party_idREG_LEAGUES -0.028586   0.037887  -0.755 0.450564
## populism       -0.000705   0.002260  -0.312 0.755143
## s(quarter)1    -1.118906   0.164407  -6.806 1.11e-11 ***
## s(quarter)2     0.435717   0.072583   6.003 2.06e-09 ***
## s(quarter)3    -0.166432   0.040374  -4.122 3.81e-05 ***
## s(quarter)4     0.175209   0.028404   6.168 7.37e-10 ***
## s(quarter)5     0.030861   0.019118   1.614 0.106534
## s(quarter)6     0.335293   0.030852  10.868 < 2e-16 ***
## s(quarter)7    -0.081241   0.052244  -1.555 0.119995
## s(quarter)8     0.647119   0.057623  11.230 < 2e-16 ***
## s(quarter)9     0.202032   0.015833  12.760 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 10:
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.135183   0.018013   7.505 7.10e-14 ***
## party_idFDI    -0.042129   0.019169  -2.198 0.028001 *
## party_idFI     0.020448   0.018402   1.111 0.266523
## party_idINDIPENDENTE 0.003060   0.034477   0.089 0.929286
## party_idIV      0.014743   0.030275   0.487 0.626307
## party_idLEGA   -0.059454   0.017469  -3.403 0.000670 ***
## party_idLEU     0.016137   0.023050   0.700 0.483902
## party_idM5S     0.075745   0.017816   4.251 2.16e-05 ***
## party_idMISTO   -0.022640   0.019159  -1.182 0.237379
## party_idPD     -0.007908   0.017773  -0.445 0.656402
## party_idREG_LEAGUES 0.051312   0.036479   1.407 0.159603
## populism       -0.004059   0.001781  -2.279 0.022683 *

```



```
## s(quarter)1      1.350315  0.145088  9.307 < 2e-16 ***
## s(quarter)2     -0.507570  0.063516 -7.991 1.61e-15 ***
## s(quarter)3      0.197423  0.036375  5.427 5.95e-08 ***
## s(quarter)4     -0.032380  0.024499 -1.322 0.186329
## s(quarter)5      0.100391  0.018161  5.528 3.39e-08 ***
## s(quarter)6     -0.208555  0.022001 -9.479 < 2e-16 ***
## s(quarter)7      0.088432  0.035241  2.509 0.012122 *
## s(quarter)8     -0.230963  0.039167 -5.897 3.92e-09 ***
## s(quarter)9     -0.043505  0.011727 -3.710 0.000209 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 11:
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.261138   0.017303  15.092 < 2e-16 ***
## party_idFDI    -0.037329   0.017620  -2.119  0.03416 *
## party_idFI      0.046592   0.016624   2.803  0.00508 **
## party_idINDIPENDENTE -0.071182  0.029682  -2.398  0.01651 *
## party_idIV      0.063320   0.031634   2.002  0.04537 *
## party_idLEGA   -0.028843   0.015837  -1.821  0.06863 .
## party_idLEU     0.012454   0.020762   0.600  0.54866
## party_idM5S    -0.014246   0.015390  -0.926  0.35463
## party_idMISTO   0.028246   0.016496   1.712  0.08690 .
## party_idPD      0.008302   0.015786   0.526  0.59896
## party_idREG_LEAGUES -0.018421  0.031182  -0.591  0.55471
## populism       0.005043   0.001929   2.615  0.00895 **
## s(quarter)1     0.419686   0.140771   2.981  0.00288 **
## s(quarter)2    -0.143947   0.060938  -2.362  0.01820 *
## s(quarter)3    -0.089103   0.033416  -2.666  0.00769 **
## s(quarter)4    -0.120523   0.021082  -5.717 1.14e-08 ***
## s(quarter)5    -0.230321   0.014344 -16.057 < 2e-16 ***
## s(quarter)6    -0.228700   0.020292 -11.270 < 2e-16 ***
## s(quarter)7    -0.216925   0.030126  -7.201 6.77e-13 ***
## s(quarter)8    -0.207864   0.034871  -5.961 2.66e-09 ***
## s(quarter)9    -0.207389   0.012606 -16.451 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Interpretation and validation

Da sistemare

```
#Topic Prevalence Over-time
#qui potresti far vedere che la prevalence di alcuni topics ha dei picchi
#in periodo in cui ha senso che abbia dei picchi...
#Anche questo vale come validation, va runnato però dopo che hai runnato estimateEffect
#perché ti servono i coefficienti dei quarters della regression...
#Anche qui, esempio con Topic 1 poi devi guardare tutti i topics...
#Va un po' sistemato (vedi x axis), magari da provare con GGLOT...
plot(prep, "quarter", method = "continuous", topics = 1,
```

```
    model = z, printlegend = FALSE, xaxt = "n", xlab = "Time")
quartseq <- seq(from = as.Date("2020-01-01"),
               to = as.Date("2022-04-18"), by = "quarter")
quartnames <- quarters(quartseq)
axis(1, at = as.numeric(quartseq) - min(as.numeric(quartseq)),
     labels = quartnames)
```