

# LDA Topicmodel

Riccardo Ruta

05/2022

## Contents

|  |          |
|--|----------|
| <b>LDA Topic model analysis</b>            | <b>1</b> |
| CREATE THE DTM . . . . .                   | 1        |
| FIND THE BEST NUMBER OF TOPICS K . . . . . | 1        |
| ANALISYS OF THE TOPICS . . . . .           | 3        |

## LDA Topic model analysis

### CREATE THE DTM

Remove all the account's mentions

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

###Convert the Document Feature Matrix (Dfm) in a Topic Model (Dtm)

```
dtm <- quanteda::convert(DFM_trimmed, to = "topicmodels")
```

### FIND THE BEST NUMBER OF TOPICS K

Search the best number of Topics comparing coherence and exclusivity values

K = 10:50

```
# 10 : 50 iter 1000
top1 <- c(10:50)

## let's create an empty data frame
risultati <- data.frame(first=vector(), second=vector(), third=vector())

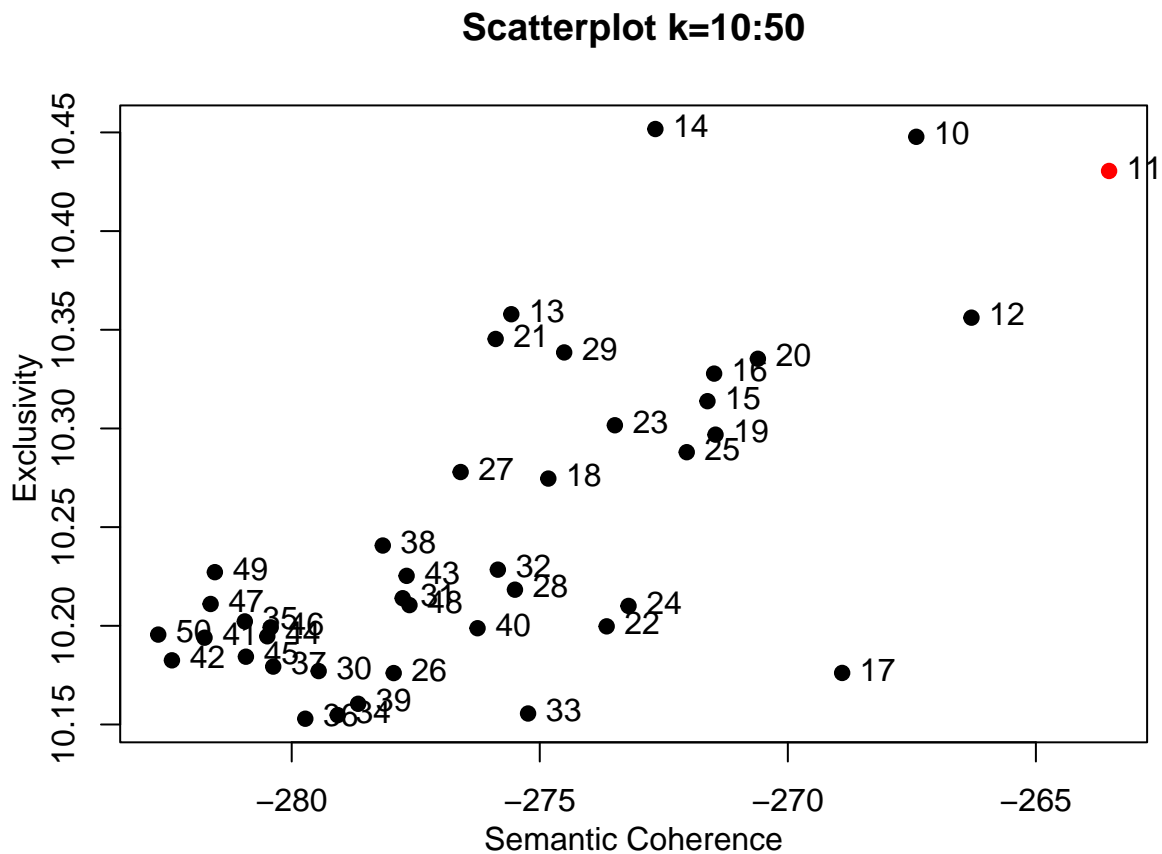
system.time(
  for (i in top1)
  {
```

```

set.seed(123)
lda_test <- LDA(dtm, method= "Gibbs", k = (i), control=list(verbose=50L, iter=1000))
topic <- (i)
coherence_test <- mean(topic_coherence(lda_test, dtm))
exclusivity_test <- mean(topic_exclusivity(lda_test))
risultati <- rbind(risultati , cbind(topic, coherence_test, exclusivity_test ))
}
)
# save(risultati,file="data/results_K_10-50.Rda")

```

Plot the values of coherence and exclusivity in order to find the best K



Interpreting those data i can state that K= 11 has the best values of coherence and exclusivity.

## ANALISYS OF THE TOPICS

Repeat the analysis selecting  $k = 11$

```
system.time(lda_11 <- LDA(dtm, method= "Gibbs", k = 11, control = list(seed = 123)))  
# save(lda_11, file = "data/lda_k_11.Rda")
```

The most important terms from the model, for each topic

| Top terms 01    | Top terms 02 | Top terms 03 | Top terms 04 | Top terms 05 | Top terms 06 |
|-----------------|--------------|--------------|--------------|--------------|--------------|
| governo         | #coronavirus | diretta      | forza_italia | solidarietà  | grande       |
| italiani        | #covid19     | roma         | bene         | libertà      | grazie       |
| lega            | sicurezza    | domani       | cose         | parole       | mondo        |
| conte           | covid        | città        | vero         | diritti      | italia       |
| matteosalvinimi | scuola       | amici        | davvero      | rispetto     | forza        |
| fratelliditalia | salute       | sindaco      | ragione      | democrazia   | de           |
| salvini         | pandemia     | insieme      | problema     | diritto      | italiano     |
| pd              | pass         | sera         | ce           | violenza     | italiana     |
| #lega           | dati         | territorio   | parla        | guerra       | storia       |
| casa            | green        | parlare      | dovrebbe     | popolo       | l'italia     |

| Top terms 7 | Top terms 8 | Top terms 9 | Top terms 10 | Top terms 11 |
|-------------|-------------|-------------|--------------|--------------|
| lavoro      | via         | governo     | anni         | piano        |
| paese       | ministro    | imprese     | grazie       | nazionale    |
| presidente  | legge       | milioni     | donne        | futuro       |
| politica    | parlamento  | euro        | giornata     | importante   |
| buon        | commissione | lavoratori  | auguri       | sociale      |
| l'italia    | senato      | famiglie    | vittime      | giovani      |
| momento     | mov5stelle  | sostegno    | famiglia     | europea      |
| bene        | voto        | cittadini   | comunità     | fondamentale |
| pdnetwork   | camera      | misure      | servizio     | paesi        |
| insieme     | intervento  | crisi       | forze        | intervista   |

```
titles_11 <- c("1", "2", "3", "4", "5", "6",  
              "7", "8", "9", "10", "11")  
  
table_titles_11 <- rbind (titles_11, terms)  
  
t11.1 <- table_titles_11[,1:6]  
t11.2 <- table_titles_11[,7:11]  
  
kable(t11.1)%>%  
  kable_styling(latex_options = c("HOLD_position", "scale_down"))
```

|           | Topic 1         | Topic 2      | Topic 3    | Topic 4      | Topic 5     | Topic 6  |
|-----------|-----------------|--------------|------------|--------------|-------------|----------|
| titles_11 | 1               | 2            | 3          | 4            | 5           | 6        |
|           | governo         | #coronavirus | diretta    | forza_italia | solidarietà | grande   |
|           | italiani        | #covid19     | roma       | bene         | libertà     | grazie   |
|           | lega            | sicurezza    | domani     | cose         | parole      | mondo    |
|           | conte           | covid        | città      | vero         | diritti     | italia   |
|           | matteosalvinimi | scuola       | amici      | davvero      | rispetto    | forza    |
|           | fratelliditalia | salute       | sindaco    | ragione      | democrazia  | de       |
|           | salvini         | pandemia     | insieme    | problema     | diritto     | italiano |
|           | pd              | pass         | sera       | ce           | violenza    | italiana |
|           | #lega           | dati         | territorio | parla        | guerra      | storia   |
|           | casa            | green        | parlare    | dovrebbe     | popolo      | l'italia |

```
kable(t11.2)%>%
  kable_styling(latex_options = c("HOLD_position","scale_down"))
```

|           | Topic 7    | Topic 8     | Topic 9    | Topic 10 | Topic 11     |
|-----------|------------|-------------|------------|----------|--------------|
| titles_11 | 7          | 8           | 9          | 10       | 11           |
|           | lavoro     | via         | governo    | anni     | piano        |
|           | paese      | ministro    | imprese    | grazie   | nazionale    |
|           | presidente | legge       | milioni    | donne    | futuro       |
|           | politica   | parlamento  | euro       | giornata | importante   |
|           | buon       | commissione | lavoratori | auguri   | sociale      |
|           | l'italia   | senato      | famiglie   | vittime  | giovani      |
|           | momento    | mov5stelle  | sostegno   | famiglia | europea      |
|           | bene       | voto        | cittadini  | comunità | fondamentale |
|           | pdnetwork  | camera      | misure     | servizio | paesi        |
|           | insieme    | intervento  | crisi      | forze    | intervista   |