Unsupervised Classification Model: Topic-model (LDA)

Miras Tolepbergen

2023-10-16

```
install.packages("topicmodels", repos='http://cran.us.r-project.org')
```

```
## Installing package into 'C:/Users/Miras/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'topicmodels' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'topicmodels'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Miras\AppData\Local\R\win-
library\4.3\00LOCK\topicmodels\libs\x64\topicmodels.dll
## to
## C:\Users\Miras\AppData\Local\R\win-
library\4.3\topicmodels\libs\x64\topicmodels.dll:
## Permission denied
```

```
## Warning: restored 'topicmodels'
```

```
##
## The downloaded binary packages are in
##  C:\Users\Miras\AppData\Local\Temp\Rtmp6Fidrk\downloaded_packages
```

```
install.packages("lubridate", repos='http://cran.us.r-project.org')
```

```
## Installing package into 'C:/Users/Miras/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'lubridate' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Miras\AppData\Local\Temp\Rtmp6Fidrk\downloaded_packages
```

```
install.packages("topicdoc", repos='http://cran.us.r-project.org')
```

```
## Installing package into 'C:/Users/Miras/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'topicdoc' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Miras\AppData\Local\Temp\Rtmp6Fidrk\downloaded_packages
```

```
install.packages("ldatuning", repos='http://cran.us.r-project.org')
```

```
## Installing package into 'C:/Users/Miras/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'ldatuning' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Miras\AppData\Local\Temp\Rtmp6Fidrk\downloaded_packages
```

```
install.packages("tidytext", repos='http://cran.us.r-project.org')
```

```
## Installing package into 'C:/Users/Miras/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'tidytext' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Miras\AppData\Local\Temp\Rtmp6Fidrk\downloaded_packages
```

```
devtools::install_github("chainsawriot/oolong")
```

```
## Skipping install of 'oolong' from a github remote, the SHA1 (b6f8aee8) has
not changed since last install.
##   Use `force = TRUE` to force installation
```

```
rm(list=ls(all=TRUE))
setwd("C:/Users/Miras/Desktop/u_m/1st/big_data_analytics/Labs/projects")
getwd()
```

```
## [1] "C:/Users/Miras/Desktop/u Milan/1st year classes/Big Data
Analystics/Labs/Lab1"
```

```
library(quanteda)
```

```
## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "replValueSp"; definition not updated

## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "xMatrix"; definition not updated

## Warning in .recacheSubclasses(def@className, def, env): undefined subclass
## "pcorMatrix" of class "mMatrix"; definition not updated

## Package version: 3.3.1
## Unicode version: 13.0
## ICU version: 69.1

## Parallel computing: 4 of 4 threads used.

## See https://quanteda.io for tutorials and examples.
```

```
library(readtext)
```

```
##
## Attaching package: 'readtext'
```

```
## The following object is masked from 'package:quanteda':
##
##     texts
```

```
library(ggplot2)
library(topicmodels)
library(tidytext)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(topicdoc)
library(cowplot)
library(ldatuning)
```

```
myText <- read.csv("guardian2013.csv", stringsAsFactors=FALSE)
str(myText)
```

```
## 'data.frame':    659 obs. of  2 variables:
##  $ date: chr  "2013-03-09" "2013-08-31" "2013-07-31" "2013-06-06" ...
##  $ text: chr  "It was the biggest US manhunt in living memory, which ended
with the fugitive taking his own life as his mounta"| __truncated__ "CANADA
\"At the present time the government of Canada has no plans, we have no plans
of our own, | to have a Ca"| __truncated__ "Alex Salmond's claims that
independence would allow Scotland to set up a better welfare system have been
suppor"| __truncated__ "The Syrian government hailed a strategic victory
yesterday after the border town of Qusair fell to Hezbollah fo"|
__truncated__ ...
```

Including Plots

```
myText$text2 <- myText$text
news_corp <- corpus(myText)
head(summary(news_corp))
```

```
##    Text Types Tokens Sentences       date
## 1 text1   245    544        23 2013-03-09
## 2 text2   158    260         6 2013-08-31
## 3 text3   135    216         6 2013-07-31
## 4 text4   386    830        28 2013-06-06
## 5 text5   393    886        34 2013-06-12
## 6 text6   144    239        12 2013-09-17
##
```

text2
1
It was the biggest US manhunt in living memory, which ended with the fugitive taking his own life as his mountain cabin hideout burned around him. But the case of Christopher Dorner, who killed four people during his rampage in southern California last month, is not over. Now comes the tussle for the reward money. | A park ranger who was carjacked by Dorner has filed a claim for the $1.2m (800,000) reward offered for information leading to his capture and arrest. | Rich Heltebrake said Dorner, dressed in military camouflage , walked up to his truck, pointed a rifle at him and said: "I don't want to hurt you, just take your dog and start walking." Heltebrake called authorities after the incident and police surrounded the cabin where Dorner was hiding and would soon die shortly after. | The reward offer was made by LA mayor Antonio Villaraigosa and other groups. "Mr Villaraigosa made a promise of that much money for the capture and conviction of Mr Dorner and I believe my phone call directly led to the end of the biggest manhunt in southern California history," Heltebrake told KTLA-5 news in LA. | In the claim filed on 19 February, his attorney wrote: "Mr Heltebrake's telephone call to Deputy Franklin notified law enforcement of Mr Dorner's location, provided a description of the vehicle he was fleeing in and was the substantial factor in the capture of Mr Dorner at the cabin location. Consequently, Mr Heltebrake accepts the mayor's offer of the entire reward of $1.2m." Dorner died of a self-inflicted gunshot wound to the head in the cabin, which was burned to the ground after police used incendiary tear gas in an attempt to drive him out. | Heltebrake said he did not think anyone expected Dorner to be captured and arrested. "But, you know, in all intents and purposes, that's what happened. When you're captured you're not free to leave," he said. "Well he was in the cabin and he wasn't free to leave." | But he is unlikely to be the sole claimant for the cash, as police received dozens of tips during the manhunt. A couple who had been tied up by Dorner in their rental cabin called 911 after freeing themselves. The carjacking of Heltebrake followed. | Dorner had also been spotted by state fish and wildlife officers who gave chase as he fled from the couple's cabin. | Heltebrake was ambiguous about whether he would split the reward. "It comes down to whether people qualify, and they had to make the claim first, that's the process, we'll see how that all goes," he said. | LA police chief Charlie Beck said authorities would decide how to distribute the money after completion of the investigation. | Captions: | Christopher Dorner killed four people during his rampage and ended up taking his own life in a mountain cabin
2
CANADA "At the present time the government of Canada has no plans, we have no plans of our own, | to have a Canadian military mission" | Prime minister Stephen Harper | FRANCE "The chemical massacre of Damascus cannot and must not remain unpunished . . . Each country is sovereign to participate or not in an operation." President Francois Hollande | UNITED STATES "History will judge us all extraordinarily harshly if we turn a blind eye to a dictator's wanton use of weapons of mass destruction" | John Kerry, secretary of state | UK "It is very clear . . . the British parliament, reflecting the views of the British people, does not want to see British military action. I get that

. . . " | Prime minister David Cameron | GERMANY "We haven't considered any German military participation and still aren't doing so" | Government spokesman Steffen Seibert | RUSSIA "People are beginning to understand how dangerous such scenarios are" Putin aide Yuri Ushakov | TURKEY "In our view there is no doubt the regime is responsible (for chemical attack)" | Foreign minister Ahmet Davutoglu | ITALY "If the United Nations does not back it, Italy will not participate" Prime minister Enrico Letta | IRAN "Western countries have found some excuse to prepare the ground to weaken the stance of Syria in further talks . . . " President Hassan Rouhani
3
Alex Salmond's claims that independence would allow Scotland to set up a better welfare system have been supported by the Institute for Fiscal Studies thinktank, in a boost for the nationalists. The IFS said an independent Scotland would be able to scrap many existing UK benefits "which make little economic sense" and discard "poorly designed reforms" by the Westminster government. | But it warned that in doing so, Scotland would probably have a higher welfare bill in future, which would cancel out recent trends where its overall benefits bill has fallen sharply from being 7% higher than the UK as a whole to get close to the UK average. In particular, Scotland's ageing population would force up costs more quickly than in the rest of the UK, unless a future Scottish government cut some benefits. | John Swinney, the Scottish finance secretary, said the IFS was correct to highlight the distorted effect that UK welfare policy had on Scottish needs. "This report confirms that people in Scotland are paying the price for high levels of housing benefit in London with a bedroom tax designed to solve a problem we simply do not have," he said.
4 The Syrian government hailed a strategic victory yesterday after the border town of Qusair fell to Hezbollah forces following a siege that pitched the Lebanese Shia militia against several thousand Sunni rebels in what had been billed as a defining battle of the country's civil war. | Rebel groups fighting the regime of President Bashar al-Assad confirmed early yesterday that they had pulled out of the town in the early hours, with fighters taking refuge in hamlets near Syria's third city, Homs, 20 miles to the north. | Outgunned since the siege began, rebels inside the town said that they had no option but to flee "in the face of this huge arsenal and lack of supplies and the blatant intervention of Hezbollah". | Opposition statements blamed "Assad regime forces aided by Iranian militias" and pledged to "continue to fight the thousands of Lebanese mercenaries". | The fate of residents who remained as the battle raged remains unclear. Rebel leaders from the town who contacted the Guardian earlier this week said more than 15,000 people had stayed in their homes from a prewar population of 30,000. | The International Committee of the Red Cross (ICRC) said it had not been given permission by Damascus to enter the town, despite winning assurances of access earlier in the week. | "We can't give any concrete information on numbers killed, wounded, or remaining in Qusair," an ICRC spokesman said. "We're still in dialogue with the Syrian authorities on reaching Qusair, particularly with a view to getting in medical supplies." | Underlining the difficulty of finding a diplomatic solution to the crisis, the UN's Syria envoy, Lakhdar Brahimi, said that a peace conference that was originally scheduled to be held in Geneva this month would now not take place before July. | Qusair had come

under heavy bombardment from artillery and shells dropped by the Syrian air force and rebel supply lines had been severed by regime forces to the north and east while Hezbollah had advanced from the south and west. | Hezbollah has led the attack. Its large-scale role has drawn strident criticism in Lebanon and across the Sunni Arab world, where inter-Muslim sectarian tensions have reached dangerous highs, especially since the assault on Qusair began. | There was no immediate reaction from the Hezbollah leadership, which sources close to the group say is anxious not to appear triumphant in Syria - in contrast to its posture after a clash with Israel. | Media outlets loyal to the group announced that the town had been "cleared of terrorists" at around 6.30am yesterday. Hezbollah is believed to have suffered close to 200 casualties during the fighting, a higher number than its members had expected before launching the attack. | Shia residents of the south Beirut suburb of Dahiyah were yesterday handing out celebratory sweets at traffic lights to mark the group's victory in Syria - a war zone far from southern Lebanon, where it has battled its traditional foe, Israel, for much of the past 30 years. | Hezbollah's role as a spearhead against a Sunni insurgency in an Arab land has forced a rethink of the group's raison d'etre and has unambiguously wedded it to the fortunes of the Assad regime. The Syrian military had been unable to gain ground in numerous battles across the country until the increased role of Hezbollah and a militia of Shia fighters from Syria and elsewhere, known as Abu Fadl al-Abbas. | Iran, the main patron of both groups, released a statement "congratulating the Syrian people for their victory". | In recent days Hezbollah had deployed hundreds of its elite forces to Qusair, a sign that the battle was drawing to a close despite resistance from rebels that had proved tougher than expected. | The defence of the town was primarily led by homegrown fighters, among them defectors. However, reinforcements from Homs and Aleppo, as well as a contingent of around 200 from the al-Qaida-aligned Jabhat al-Nusra, arrived one week ago. The total number of defenders is thought to have numbered around 3,000. | Rebel casualty figures are unknown. But on Monday, a surgeon from the town contacted the Guardian to say his supplies of vital medicines had run out. "Nothing has been able to get through, he said." | Captions: | Syrian armour outside Qusair, but Hezbollah played a decisive role | Some residents, such as this child, escaped to Lebanon as the town fell | The fate of those who endured the three-week siege is unknown

5

He may be in his 90th year but "Mad" Frankie Fraser is still causing mayhem. It has emerged that the former gangland enforcer, who has spent 42 years in prison for 26 offences, has been issued with an asbo after an incident in his residential accommodation. | Fraser, who was jailed for 10 years in the so-called "torture trial" in 1967, is now frail and in poor health. However, according to a new documentary, he is clearly not going gentle into any good night. He was given an asbo, one of his sons told film-makers, after getting into an argument with a fellow-resident and is unrepentant about his life of crime. | Last seen in public in October at the funeral of his former boss, Charlie Richardson, Fraser is one of the few remaining members of a generation of "celebrity criminals". His life of crime started aged nine when he worked for the notorious Sabini gang, which ran protection rackets at the racecourses at a time when off-course betting was illegal. | "At the races,

I'd be bucket boy," says Fraser in the documentary, Frankie Fraser's Last Stand, which will be broadcast on the Crime and Investigation network on 16 June at 9pm. A bucket boy would offer to clean the bookies' blackboards with a sponge, for which they were obliged to pay the Sabinis. Both Fraser and his sister, Eva, were also active juvenile thieves. "You name it, we nicked it," he says. "As I was growing up, I never had to buy a shirt - Eva made sure she nicked them for me." | A deserter during the war - he pretended to be mad to avoid the call-up - Fraser was certified insane three times and spent time in Broadmoor secure hospital. He was frequently punished for breaking prison rules or fighting prison officers: "I've done more bread and water than any man alive." | Of the war years, when he was heavily involved in theft from bombed-out stores, he says: "You wanted to win the war but you wanted it to go on for ever. It was a thief's paradise, Gor blimey! Whatever you nicked you could sell, they'd be queuing up to buy it off you." | After the war, he worked for underworld boss Billy Hill, for whom he carried out razor attacks. "Hill paid by the stitch - if you put 50 stitches in a man's face, you could expect 50," says James Morton, Fraser's biographer. News reports were checked to see how much was owing. | Fraser was jailed along with other members of the Richardson gang for violently punishing people whom the Richardsons believed owed them money. His decision to join the Richardsons rather than their rivals, the Krays, has been described as "like China getting the atom bomb". | He emerged from jail in 1989 and has not been back since. "Maybe he was bored with going to prison," Ronnie Richardson, Charlie's widow, tells the programme. In 1991, while emerging from Turnmills nightclub in Clerkenwell, London, he was shot at by an unidentified gunman. According to one of his sons, David, Fraser was unharmed but he did not inform on his assailant. "If you play by the sword, you've got to expect the sword as well," says his son. | Over the last decade or so he was on the cabaret circuit and ran gangland tours of the East End, taking in such sights as the Blind Beggar pub, where Ronnie Kray shot dead George Cornell, one of the Richardson gang, in 1966. | A famous Monty Python sketch featuring the Piranha brothers, Doug and Dinsdale, has often been associated with Fraser and the Kray twins and some aspects of the new documentary may add to this impression. | Fraser has complained in the past that "I had no help from my family; my mother and father were dead straight so I had to make my own way." | The new documentary returns to this theme, suggesting he had a hard time in prison because there were no criminals in his family. "My father was the most honest man I've ever come across," says Fraser, who also refers to his Native American antecedents, saying that his grandmother was "a Red Indian". According to his sons, Fraser has no regrets: "He said, 'No, I wouldn't have done my life any other way.'" | Captions: | Former gangland enforcer Frankie Fraser has no regrets about his life of crime, despite having spent 42 years behind bars for 26 offences

6

A project to reintroduce a bumblebee unseen in the UK for 25 years is celebrating after experts confirmed queens have nested and produced young. | Over two years conservationists have made two releases of short-haired bumblebee (Bombus subterraneus) queens at the RSPB reserve in Dungeness, Kent, in an attempt to establish a viable breeding population. They have recorded seven young worker bees in a 4 sq km area. | Dr Nikki Gammans, who

```
leads the project, said: "This is a milestone . . . and a real victory for
conservation. We now have proof this bumblebee has nested and hatched young,
and we hope it is on the way to becoming a self-supporting wild species in
the UK once again." | The short-haired bumblebee has not been seen in the UK
since 1988, and was declared extinct in 2000. The species was once widespread
across the south of England, feeding on wildflowers such as white dead-nettle
and red clover. But it declined in the second half of the 20th century as
wildflower-rich grassland was lost to intensively farmed land. | Previous
efforts to reintroduce the bee have fared poorly. In 2009 scientists tried to
transport queens from New Zealand. But the colony lacked genetic diversity,
and many of the queens died in quarantine. Jessica Aldred
```

```
tok <- tokens(news_corp, remove_punct = TRUE, remove_numbers=TRUE,
remove_symbols = TRUE, split_hyphens = TRUE, remove_separators = TRUE)
tok <- tokens_remove(tok, stopwords("en"))
tok <- tokens_wordstem (tok)
news_dfm <- dfm(tok)
news_dfm <- dfm_remove(news_dfm, c('*-time', '*-timeUpdated', 'GMT', 'BST'))
news_dfm <-   dfm_trim(news_dfm, min_termfreq = 0.95, termfreq_type =
"quantile",
                       max_docfreq = 0.1, docfreq_type = "prop")
news_dfm[1:2, 1:5]
```

```
## Document-feature matrix of: 2 documents, 5 features (40.00% sparse) and 2
docvars.
##        features
## docs    park arrest militari incid mr
##   text1    1      2        1     1  6
##   text2    0      0        3     0  0
```

```
str(docvars(news_dfm))
```

```
## 'data.frame':    659 obs. of  2 variables:
##  $ date : chr  "2013-03-09" "2013-08-31" "2013-07-31" "2013-06-06" ...
##  $ text2: chr  "It was the biggest US manhunt in living memory, which
ended with the fugitive taking his own life as his mounta"| __truncated__
"CANADA \"At the present time the government of Canada has no plans, we have
no plans of our own, | to have a Ca"| __truncated__ "Alex Salmond's claims
that independence would allow Scotland to set up a better welfare system have
been suppor"| __truncated__ "The Syrian government hailed a strategic victory
yesterday after the border town of Qusair fell to Hezbollah fo"|
__truncated__ ...
```

```
topfeatures(news_dfm, 20)
```

```
##      tax  council   custom     food militari     vote       co     data
##      210      177      162      158      148      146      135      132
##   health     book      war    north  cameron   prison    board     test
##      130      130      129      127      126      125      124      123
##   invest    china  protest    trust
##      123      122      122      120
```

```
table(ntoken(news_dfm)== 0)
```

```
##
## FALSE  TRUE
##   658     1
```

```
news_dfm <- news_dfm[ntoken(news_dfm) > 0,]
```

```
dtm <- convert(news_dfm, to = "topicmodels")
```

Identifying the optimal number of topics: coherence and exclusivity

```
top <- c(4:25) # let's change k between 4 and 25 and each time we store the
corresponding avg. values of both coherence and exclusivity
top
```

```
##  [1]  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
```

```
results <- data.frame(first=vector(), second=vector(), third=vector())
# let's create an empty data frame that we will fill later on
results
```

```
## [1] first  second third
## <0 rows> (or 0-length row.names)
```

```
system.time(
for (i  in top)
{
set.seed(123)
lda <- LDA(dtm, method= "Gibbs", k = i,  control=list(verbose=50L,
iter=1000))
topic <- i
coherence <- mean(topic_coherence(lda, dtm))
exclusivity <- mean(topic_exclusivity(lda))
results <- rbind(results , cbind(topic, coherence, exclusivity ))
    }
)
```

```
## K = 4; V = 366; M = 658
## Sampling 1000 iterations!
## Iteration 50 ...
```

```
##    user  system elapsed
##   56.37    0.14   58.18
```

```
results
```

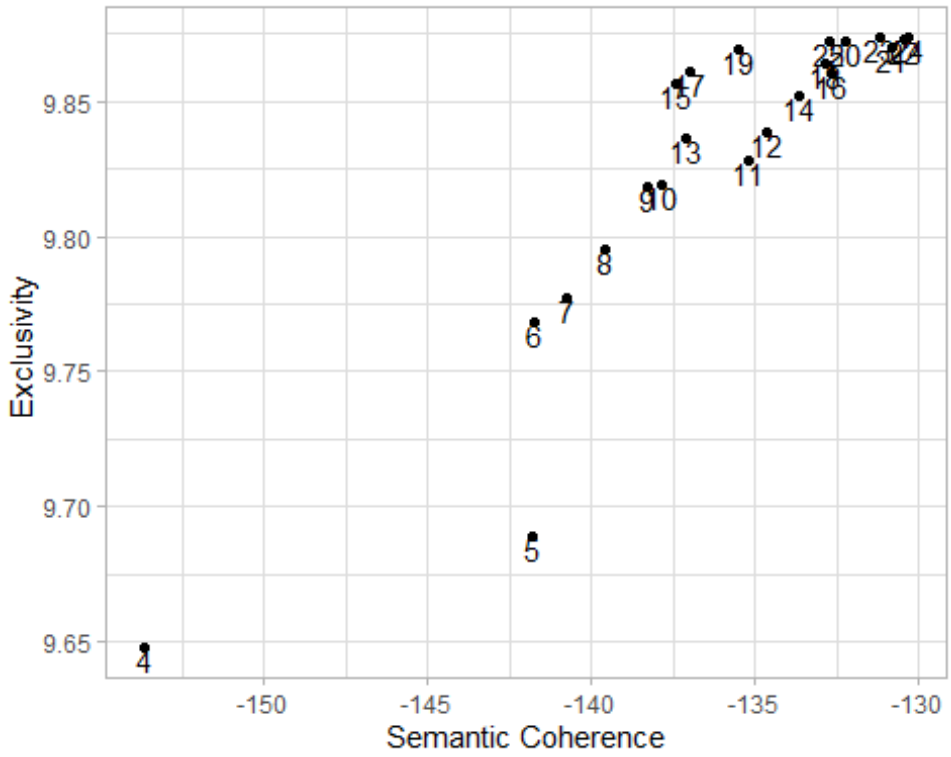
```
##    topic coherence exclusivity
## 1      4 -153.6605    9.648049
## 2      5 -141.8283    9.689108
## 3      6 -141.7060    9.768270
```

```
## 4      7 -140.7492    9.777123
## 5      8 -139.5725    9.795248
## 6      9 -138.2462    9.818362
## 7     10 -137.8260    9.819080
## 8     11 -135.1654    9.827842
## 9     12 -134.6157    9.838527
## 10    13 -137.0915    9.836065
## 11    14 -133.6376    9.851918
## 12    15 -137.4316    9.856237
## 13    16 -132.6849    9.860232
## 14    17 -137.0070    9.860995
## 15    18 -132.8257    9.864000
## 16    19 -135.5221    9.869332
## 17    20 -132.2217    9.872495
## 18    21 -130.8115    9.870240
## 19    22 -130.4130    9.872814
## 20    23 -131.1713    9.873338
## 21    24 -130.2914    9.873745
## 22    25 -132.7230    9.872530

str(results)

## 'data.frame':    22 obs. of  3 variables:
##  $ topic      : num  4 5 6 7 8 9 10 11 12 13 ...
##  $ coherence  : num  -154 -142 -142 -141 -140 ...
##  $ exclusivity: num  9.65 9.69 9.77 9.78 9.8 ...

ggplot(results, aes(x=coherence, y=exclusivity)) + geom_point() +
  geom_text(label=results$topic, vjust=1) +
ylab(label="Exclusivity ") +  xlab("Semantic Coherence")  +
  theme_light()
```



#let's try k=24 topics

```
set.seed(123)
system.time(lda <- LDA(dtm, method= "Gibbs", k = 24,
control=list(verbose=50L)))

## K = 24; V = 366; M = 658
## Sampling 2000 iterations!
## Iteration 50 ...
## Iteration 100 ...
## Iteration 150 ...
## Iteration 200 ...
## Iteration 250 ...
## Iteration 300 ...
## Iteration 350 ...
## Iteration 400 ...
## Iteration 450 ...
## Iteration 500 ...
## Iteration 550 ...
## Iteration 600 ...
## Iteration 650 ...
## Iteration 700 ...
## Iteration 750 ...
## Iteration 800 ...
## Iteration 850 ...
## Iteration 900 ...
## Iteration 950 ...
```

```
## Iteration 1000 ...
## Iteration 1050 ...
## Iteration 1100 ...
## Iteration 1150 ...
## Iteration 1200 ...
## Iteration 1250 ...
## Iteration 1300 ...
## Iteration 1350 ...
## Iteration 1400 ...
## Iteration 1450 ...
## Iteration 1500 ...
## Iteration 1550 ...
## Iteration 1600 ...
## Iteration 1650 ...
## Iteration 1700 ...
## Iteration 1750 ...
## Iteration 1800 ...
## Iteration 1850 ...
## Iteration 1900 ...
## Iteration 1950 ...
## Iteration 2000 ...
## Gibbs sampling completed!

##    user  system elapsed
##    7.80    0.03    8.24

termsList <- get_terms(lda, 10)
terms(lda, 10)  #topics look mutually exclusive

##       Topic 1    Topic 2   Topic 3  Topic 4     Topic 5   Topic 6    Topic
7
##  [1,] "invest"   "travel"  "china"  "test"      "design"  "committe" "co"
##  [2,] "profit"   "train"   "site"   "men"       "film"    "lord"
"cover"
##  [3,] "trade"    "minut"   "church" "list"      "seri"    "event"
"regul"
##  [4,] "loss"     "station" "visit"  "cultur"    "music"   "fine"     "bbc"
##  [5,] "annual"   "arriv"   "poll"   "ban"       "idea"    "discuss"  "op"
##  [6,] "institut" "park"    "french" "februari"  "space"   "sir"
"insur"
##  [7,] "drop"     "wait"    "chines" "communiti" "tv"      "prove"    "bad"
##  [8,] "investor" "began"   "war"    "recommend" "award"   "lack"
"debt"
##  [9,] "corpor"   "saw"     "franc"  "respond"   "watch"   "brought"
"condit"
## [10,] "sell"     "someon"  "word"   "cours"     "project" "reach"
"reduc"
##       Topic 8    Topic 9      Topic 10    Topic 11   Topic 12   Topic 13
##  [1,] "book"     "tax"        "obama"     "food"     "north"    "european"
##  [2,] "phone"    "budget"     "american"  "chariti"  "defenc"   "growth"
```

```
##  [3,] "access"   "pension"    "intellig"  "aid"      "militari" "euro"
##  [4,] "apple"    "osborn"     "collect"   "standard" "town"     "europ"
##  [5,] "internet" "earn"       "document"  "poverti"  "al"       "global"
##  [6,] "email"    "georg"      "secret"    "drug"     "regim"    "quarter"
##  [7,] "letter"   "chancellor" "administr" "suppli"   "weapon"   "eu"
##  [8,] "noth"     "incom"      "agenc"     "safeti"   "syria"    "recoveri"
##  [9,] "messag"   "financ"     "special"   "estim"    "fight"
"manufactur"
## [10,] "read"     "reform"     "repres"    "crisi"    "war"      "septemb"
##       Topic 14     Topic 15  Topic 16   Topic 17       Topic 18    Topic
19
##  [1,] "vote"       "trust"   "protest"  "board"        "council"
"employ"
##  [2,] "cameron"    "mandela" "arm"      "heard"        "scheme"
"contract"
##  [3,] "conserv"    "health"  "fire"     "team"         "averag"    "royal"
##  [4,] "scotland"   "hospit"  "morsi"    "relationship" "properti"
"worker"
##  [5,] "mps"        "nhs"     "violenc"  "understand"   "mortgag"   "union"
##  [6,] "tori"       "mr"      "muslim"   "bodi"         "april"     "lost"
##  [7,] "coalit"     "doctor"  "student"  "sector"       "owner"     "worri"
##  [8,] "speech"     "seven"   "activist" "data"         "household"
"agreement"
##  [9,] "referendum" "send"    "militari" "hear"         "low"
"office"
## [10,] "democrat"   "medic"   "thousand" "approach"     "survey"    "lose"
##       Topic 20    Topic 21  Topic 22    Topic 23   Topic 24
##  [1,] "advis"     "woman"   "custom"    "water"    "prison"
##  [2,] "lot"       "arrest"  "retail"    "research" "black"
##  [3,] "quit"      "child"   "store"     "studi"    "justic"
##  [4,] "career"    "mother"  "shop"      "energi"   "trial"
##  [5,] "interview" "parent"  "onlin"     "self"     "judg"
##  [6,] "matter"    "inquiri" "complaint" "popul"    "human"
##  [7,] "anyth"     "stay"    "data"      "affect"   "abus"
##  [8,] "let"       "son"     "payment"   "generat"  "crimin"
##  [9,] "certain"   "girl"    "cash"      "born"     "victim"
## [10,] "progress"  "murder"  "card"      "poor"     "crime"

lda_topics <- tidy(lda, matrix = "beta")
str(lda_topics)

## tibble [8,784 × 3] (S3: tbl_df/tbl/data.frame)
##  $ topic: int [1:8784] 1 2 3 4 5 6 7 8 9 10 ...
##  $ term : chr [1:8784] "park" "park" "park" "park" ...
##  $ beta : num [1:8784] 8.78e-05 4.90e-02 9.64e-05 9.90e-05 8.55e-05 ...

top_terms <- group_by(lda_topics, topic)
str(top_terms)

## gropd_df [8,784 × 3] (S3: grouped_df/tbl_df/tbl/data.frame)
##  $ topic: int [1:8784] 1 2 3 4 5 6 7 8 9 10 ...
```

```
##  $ term : chr [1:8784] "park" "park" "park" "park" ...
##  $ beta : num [1:8784] 8.78e-05 4.90e-02 9.64e-05 9.90e-05 8.55e-05 ...
##  - attr(*, "groups")= tibble [24 × 2] (S3: tbl_df/tbl/data.frame)
##   ..$ topic: int [1:24] 1 2 3 4 5 6 7 8 9 10 ...
##   ..$ .rows: list<int> [1:24]
##   .. ..$ : int [1:366] 1 25 49 73 97 121 145 169 193 217 ...
##   .. ..$ : int [1:366] 2 26 50 74 98 122 146 170 194 218 ...
##   .. ..$ : int [1:366] 3 27 51 75 99 123 147 171 195 219 ...
##   .. ..$ : int [1:366] 4 28 52 76 100 124 148 172 196 220 ...
##   .. ..$ : int [1:366] 5 29 53 77 101 125 149 173 197 221 ...
##   .. ..$ : int [1:366] 6 30 54 78 102 126 150 174 198 222 ...
##   .. ..$ : int [1:366] 7 31 55 79 103 127 151 175 199 223 ...
##   .. ..$ : int [1:366] 8 32 56 80 104 128 152 176 200 224 ...
##   .. ..$ : int [1:366] 9 33 57 81 105 129 153 177 201 225 ...
##   .. ..$ : int [1:366] 10 34 58 82 106 130 154 178 202 226 ...
##   .. ..$ : int [1:366] 11 35 59 83 107 131 155 179 203 227 ...
##   .. ..$ : int [1:366] 12 36 60 84 108 132 156 180 204 228 ...
##   .. ..$ : int [1:366] 13 37 61 85 109 133 157 181 205 229 ...
##   .. ..$ : int [1:366] 14 38 62 86 110 134 158 182 206 230 ...
##   .. ..$ : int [1:366] 15 39 63 87 111 135 159 183 207 231 ...
##   .. ..$ : int [1:366] 16 40 64 88 112 136 160 184 208 232 ...
##   .. ..$ : int [1:366] 17 41 65 89 113 137 161 185 209 233 ...
##   .. ..$ : int [1:366] 18 42 66 90 114 138 162 186 210 234 ...
##   .. ..$ : int [1:366] 19 43 67 91 115 139 163 187 211 235 ...
##   .. ..$ : int [1:366] 20 44 68 92 116 140 164 188 212 236 ...
##   .. ..$ : int [1:366] 21 45 69 93 117 141 165 189 213 237 ...
##   .. ..$ : int [1:366] 22 46 70 94 118 142 166 190 214 238 ...
##   .. ..$ : int [1:366] 23 47 71 95 119 143 167 191 215 239 ...
##   .. ..$ : int [1:366] 24 48 72 96 120 144 168 192 216 240 ...
##   .. ..@ ptype: int(0)
##   ..- attr(*, ".drop")= logi TRUE

# let's keep only the first top 4 betas for each of the topics
top_terms <- top_n(top_terms, 4, beta)
top_terms <- ungroup(top_terms)
top_terms <- arrange(top_terms , topic, -beta)
str(top_terms)

## tibble [97 × 3] (S3: tbl_df/tbl/data.frame)
##  $ topic: int [1:97] 1 1 1 1 2 2 2 2 3 3 ...
##  $ term : chr [1:97] "invest" "profit" "trade" "loss" ...
##  $ beta : num [1:97] 0.1019 0.0835 0.0773 0.0606 0.082 ...

table(top_terms$topic)

##
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
##  4  4  4  4  4  4  4  4  4  4  4  4  4  5  4  4  4  4  4  4  4  4  4  4
```

```
top_terms <-  mutate(top_terms,   topic = factor(topic),
    term = reorder_within(term, beta, topic))
str(top_terms)
```

```
## tibble [97 × 3] (S3: tbl_df/tbl/data.frame)
##  $ topic: Factor w/ 24 levels "1","2","3","4",..: 1 1 1 1 2 2 2 2 3 3 ...
##  $ term : Factor w/ 97 levels "career___20",..: 88 70 59 26 67 66 25 20 92
69 ...
##   ..- attr(*, "scores")= num [1:97(1d)] 0.0658 0.0571 0.0568 0.061 0.0526
...
##   .. ..- attr(*, "dimnames")=List of 1
##   .. .. ..$ : chr [1:97] "access___8" "advis___20" "aid___11"
"american___10" ...
##  $ beta : num [1:97] 0.1019 0.0835 0.0773 0.0606 0.082 ...
```

```
ggplot(top_terms, aes(term, beta, fill = topic)) +
geom_bar(alpha = 0.8, stat = "identity", show.legend = FALSE) +
scale_x_reordered() +
facet_wrap(facets = vars(topic), scales = "free", ncol = 5)+
coord_flip()
```


```
#there are no overlaps of terms between the topics which indicates a high
level of semantic coherence. Let's validate though

head(topics(lda))
```

```
## text1 text2 text3 text4 text5 text6
##    15    12    14    12    24     5
```

```r
str(docvars(news_dfm))
```

```
## 'data.frame':    658 obs. of  2 variables:
##  $ date : chr  "2013-03-09" "2013-08-31" "2013-07-31" "2013-06-06" ...
##  $ text2: chr  "It was the biggest US manhunt in living memory, which
ended with the fugitive taking his own life as his mounta"| __truncated__
"CANADA \"At the present time the government of Canada has no plans, we have
no plans of our own, | to have a Ca"| __truncated__ "Alex Salmond's claims
that independence would allow Scotland to set up a better welfare system have
been suppor"| __truncated__ "The Syrian government hailed a strategic victory
yesterday after the border town of Qusair fell to Hezbollah fo"|
__truncated__ ...
```

```r
docvars(news_dfm, 'pred_topic') <- topics(lda)
str(docvars(news_dfm))
```

```
## 'data.frame':    658 obs. of  3 variables:
##  $ date      : chr  "2013-03-09" "2013-08-31" "2013-07-31" "2013-06-06"
...
##  $ text2     : chr  "It was the biggest US manhunt in living memory, which
ended with the fugitive taking his own life as his mounta"| __truncated__
"CANADA \"At the present time the government of Canada has no plans, we have
no plans of our own, | to have a Ca"| __truncated__ "Alex Salmond's claims
that independence would allow Scotland to set up a better welfare system have
been suppor"| __truncated__ "The Syrian government hailed a strategic victory
yesterday after the border town of Qusair fell to Hezbollah fo"|
__truncated__ ...
##  $ pred_topic: int  15 12 14 12 24 5 21 2 5 24 ...
```

```r
head(lda@gamma)
```

```
##           [,1]       [,2]       [,3]       [,4]       [,5]       [,6]
## [1,] 0.02815315 0.05518018 0.04166667 0.04166667 0.04166667 0.02815315
## [2,] 0.03255208 0.03255208 0.03255208 0.03255208 0.03255208 0.03255208
## [3,] 0.04282407 0.02893519 0.04282407 0.02893519 0.07060185 0.02893519
## [4,] 0.03490028 0.03490028 0.01780627 0.01780627 0.01780627 0.06054131
## [5,] 0.03146259 0.02125850 0.03146259 0.03146259 0.05187075 0.04166667
## [6,] 0.03109453 0.04601990 0.03109453 0.04601990 0.07587065 0.04601990
##           [,7]       [,8]       [,9]      [,10]      [,11]      [,12]
## [1,] 0.02815315 0.04166667 0.02815315 0.04166667 0.04166667 0.04166667
## [2,] 0.04817708 0.04817708 0.03255208 0.04817708 0.04817708 0.11067708
## [3,] 0.04282407 0.04282407 0.05671296 0.02893519 0.04282407 0.02893519
## [4,] 0.02635328 0.04344729 0.01780627 0.02635328 0.06908832 0.35968661
## [5,] 0.03146259 0.04166667 0.03146259 0.03146259 0.02125850 0.04166667
## [6,] 0.03109453 0.03109453 0.03109453 0.04601990 0.04601990 0.03109453
##          [,13]      [,14]      [,15]      [,16]      [,17]      [,18]
## [1,] 0.02815315 0.02815315 0.14977477 0.04166667 0.02815315 0.02815315
## [2,] 0.04817708 0.06380208 0.03255208 0.03255208 0.03255208 0.03255208
## [3,] 0.02893519 0.08449074 0.02893519 0.02893519 0.07060185 0.05671296
## [4,] 0.01780627 0.01780627 0.01780627 0.03490028 0.01780627 0.01780627
## [5,] 0.02125850 0.02125850 0.04166667 0.02125850 0.02125850 0.02125850
```

```
## [6,] 0.03109453 0.04601990 0.06094527 0.03109453 0.03109453 0.03109453
##           [,19]      [,20]      [,21]      [,22]      [,23]      [,24]
## [1,] 0.02815315 0.02815315 0.06869369 0.02815315 0.04166667 0.04166667
## [2,] 0.03255208 0.03255208 0.04817708 0.03255208 0.03255208 0.04817708
## [3,] 0.02893519 0.02893519 0.02893519 0.02893519 0.07060185 0.02893519
## [4,] 0.02635328 0.02635328 0.03490028 0.01780627 0.02635328 0.01780627
## [5,] 0.02125850 0.03146259 0.08248299 0.04166667 0.08248299 0.18452381
## [6,] 0.06094527 0.04601990 0.04601990 0.03109453 0.06094527 0.03109453

round(lda@gamma[1,], 2)

##  [1] 0.03 0.06 0.04 0.04 0.04 0.03 0.03 0.04 0.03 0.04 0.04 0.04 0.03 0.03
0.15
## [16] 0.04 0.03 0.03 0.03 0.03 0.07 0.03 0.04 0.04

max(lda@gamma[,1])#maximum value of theta for topic 1 in the dataframe

## [1] 0.2508333

which.max(lda@gamma[,1])

## [1] 433

strwrap(news_dfm@docvars$text2[433])

##  [1] "European technology companies are trapped in a funding bottleneck
as"
##  [2] "venture capital firms refuse to reinvest in the sector until they
see a"
##  [3] "return from earlier investments, such as music service Spotify, via"
##  [4] "successful stockmarket flotations. | As US companies such as
Twitter"
##  [5] "gear up for multibillion-dollar flotations, European venture"
##  [6] "capitalists say that they need to see share issues that will recoup"
##  [7] "their initial forays into the technology sector before seeking out
the"
##  [8] "next Google. | One senior investment banker said success on the
London"
##  [9] "Stock Exchange (LSE) was essential for companies such as Spotify,"
## [10] "property website Zoopla and Klarna, a mobile payment service. Funds"
## [11] "raised by a wave of initial public offerings in technology
companies"
## [12] "can then be returned to venture capitalists, who typically invest
small"
## [13] "sums in startup businesses and then reap the reward when they float
or"
## [14] "become a commercial success. | \"We need some good visible success"
## [15] "stories, floating on major European markets - really the LSE,\" the"
## [16] "banker said. \"These companies need to perform well in the after-
market"
## [17] "and make money for the institutional investors who buy in the IPO"
## [18] "(initial public offering) - that way we will start to build a"
```

```
## [19] "sustainable environment for financing, building and exiting
(selling"
## [20] "stakes in) tech companies in Europe.\" | The banker said once
investors"
## [21] "saw a return from technology investments in the wake of flotations,"
## [22] "more funds would become available. \"This will eventually lead top-
tier"
## [23] "venture capital firms to be able to sell down their stake so they
can"
## [24] "return money to their investors. If the flotations go well there is
an"
## [25] "expectation that the UK market will become more receptive to"
## [26] "venture-backed tech companies.\" | He added: \"We believe that for
the"
## [27] "LSE you need to focus on demonstrating that the companies are great"
## [28] "companies in their own right, not 'tech companies' - for instance
Asos"
## [29] "and Rightmove.\" | Asos, an online clothing retailer, joined the"
## [30] "Alternative Investment Market (AIM) in October 2001, just over a
year"
## [31] "after setting up, and is now valued at 4bn, while property website"
## [32] "Rightmove floated in 2006 and is now valued at 2.4bn. | Although"
## [33] "technology startups in the UK are benefiting from a growing pool of"
## [34] "\"angel\" investors, who put in small amounts, typically a few
thousand"
## [35] "pounds, to kickstart a company, many are struggling to secure their"
## [36] "first institutional investment rounds, know as Series A and B. |"
## [37] "Industry sources say this is because many venture capital funds
that"
## [38] "have seen low returns from past Series A and B investments are
turning"
## [39] "to late-stage investment when companies are bigger, depriving funds
for"
## [40] "medium-sized firms. | Spotify raised $100m ( 64.2m) from investors
in"
## [41] "September 2012, in a move that valued the business at $3bn. But the"
## [42] "Swedish music-streaming company has not given an indication that it"
## [43] "intends to float on the stock market before 2015. UK-based Zoopla
is"
## [44] "seen as a rival to Rightmove, and is expected to generate interest
from"
## [45] "institutional investors."

round(lda@gamma[1,], 2)

##  [1] 0.03 0.06 0.04 0.04 0.04 0.03 0.03 0.04 0.03 0.04 0.04 0.04 0.03 0.03
0.15
## [16] 0.04 0.03 0.03 0.03 0.03 0.07 0.03 0.04 0.04
```

```r
# an article with the second highest gamma(=theta) for topic 1
sort(lda@gamma[,1])[length(lda@gamma[,1]) - 1]
```

```
## [1] 0.2437695
```

```r
which(lda@gamma[,1]==sort(lda@gamma[,1])[length(lda@gamma[,1]) - 1]  )
```

```
## [1] 353
```

```r
strwrap(news_dfm@docvars$text2[353])
```

```
##  [1] "It's the 75,000 investment that has not only caused the archbishop
of"
##  [2] "Canterbury intense embarrassment but also shone a spotlight on some
of"
##  [3] "the Church of England's other ethically questionable investments. |"
##  [4] "Less than 24 hours after the Most Rev Justin Welby said he wanted
to"
##  [5] "\"compete Wonga out of existence\" it was revealed the Church of
England"
##  [6] "holds an indirect investment in the payday lender. Yesterday Welby"
##  [7] "promised an inquiry to discover how the investment slipped past the"
##  [8] "church's ethical investment advisory group. | But it later emerged
that"
##  [9] "the investment does not breach the church's ethical investment
policy,"
## [10] "which only prevents the church from directly investing in companies"
## [11] "that make more than 25% of their money from high-interest lending.
In"
## [12] "this case the church commissioners' 5.5bn fund was not directly"
## [13] "invested in Wonga, but via a \"fund of funds\" called VenCap, which
holds"
## [14] "an investment in Accel Partners, one of Wonga's main financial
backers."
## [15] "| Gavin Oldham, a member of the church commissioner asset committee
and"
## [16] "ethical investment advisory group, said the investment represents
\"a"
## [17] "tiny fraction of percent\" of the 5.5bn fund. | Oldman said the
church's"
## [18] "investment managers were aware of the Wonga investment but had not"
## [19] "passed it on to Welby or the church commissioners, which he
admitted"
## [20] "was \"very embarrassing\". | But Oldman said the investment was of
an"
## [21] "\"immaterial level\" and could be compared to investing in Tesco,
which"
## [22] "\"sells topshelf pornography\". The church has also banned itself
from"
## [23] "investing directly in companies that make a significant amount of
money"
```

```
## [24] "from pornography, tobacco and arms dealing. | Conservative MP
Claire"
## [25] "Perry, an adviser to prime minister David Cameron on preventing the"
## [26] "sexualisation and commercialisation of childhood, urged the church
to"
## [27] "consider pulling its money out of Google to increase pressure on
the"
## [28] "company to block child abuse images. She told the Daily Telegraph:"
## [29] "\"They (the Church of England and other investors) have a role to
play,"
## [30] "they have questions to ask themselves. They are moral leaders. If
they"
## [31] "are going to opine on things then putting your money where your
mouth"
## [32] "is is an incredibly powerful tool.\" | The church commissioners'
latest"
## [33] "annual report shows the church's top five corporate investments -
Royal"
## [34] "Dutch Shell, HSBC, BP, Vodafone and GlakoSmithKline - have all been"
## [35] "embroiled in recent scandals. | Oldman said it was \"completely"
## [36] "unrealistic\" to expect the church to disinvest from big British"
## [37] "companies despite unethical behaviour. \"It's tricky, isn't it,\"
he"
## [38] "said. \"The whole world is controversial.\" | Captions: | The
archbishop"
## [39] "said he wanted to 'compete Wonga out of business'"
```

```
# an article with the third highest gamma(=theta) for topic 1
sort(lda@gamma[,1])[length(lda@gamma[,1]) - 2]
```

```
## [1] 0.2420977
```

```
which(lda@gamma[,1]==sort(lda@gamma[,1])[length(lda@gamma[,1]) - 2]  )
```

```
## [1] 122
```

```
strwrap(news_dfm@docvars$text2[122])
```

```
##  [1] "John Malone's takeover of Virgin Media will create a $28bn ( 18bn)"
##  [2] "company headquartered in Britain, but the world's largest cable
group"
##  [3] "by customers will pay no UK tax for the foreseeable future. |
Malone's"
##  [4] "Liberty Global, which owns 11 cable companies in Europe, has
confirmed"
##  [5] "an agreed cash and shares bid that allows Virgin chief executive
Neil"
##  [6] "Berkett to walk away with shares worth $78m and gives 2,500 fellow"
##  [7] "staff nearly 16,000 each from an employee share scheme. | The
combined"
##  [8] "company, which will retain the Virgin brand in Britain but not its"
```

```
##  [9] "boss, would have more customers than America's largest group,
Comcast,"
## [10] "and the muscle to challenge Rupert Murdoch's UK dominance of pay TV.
|"
## [11] "But the losses accumulated by Virgin Media after two decades of"
## [12] "investment in the country's cable network mean the new company would
be"
## [13] "exempt from tax for at least 15 years, according to analysts at"
## [14] "Espirito Santo bank. Bernstein bank said Liberty would pay no UK
taxes"
## [15] "for \"the foreseeable future\". | Virgin was formed in 2006 from
the"
## [16] "alliance of Virgin Mobile, NTL and Telewest, which were assembled
by"
## [17] "merging regional cable networks. After investing 13bn in laying
fibre"
## [18] "optic cables to half of UK homes, the industry was loss-making
until"
## [19] "Virgin began to turn a profit in 2010. | Because it has been
profitable"
## [20] "for three years, it can declare how much of its historic losses it
will"
## [21] "offset against profits. Yesterday, Virgin put that number at 2.6bn -
"
## [22] "allowing significant leeway on its balance sheet before it needs to
pay"
## [23] "corporation tax in Britain. In effect, Virgin needs to accumulate"
## [24] "profits far in excess of 2.6bn before it starts paying corporation
tax."
## [25] "The company made 261m in pre-tax profit in 2012, which is forecast
to"
## [26] "rise to 500m after 2014. On that basis, Virgin would have been
paying"
## [27] "100m a year into the public purse if it was a regular payer of"
## [28] "corporation tax at a rate of 21% of profits. | \"You're looking at"
## [29] "probably 15 years for them to work their way through, but this is a"
## [30] "company that over the last 20 years has invested billions in
network"
## [31] "infrastructure,\" said Espirito analyst Andrew Hogley. \"This is a
very"
## [32] "different situation to a Starbucks or an Amazon, that is offshoring"
## [33] "profits to avoid paying UK taxes.\" | Malone's entry into the UK
cable"
## [34] "market marks the culmination of a decade long ambition. In 2002,
the"
## [35] "man known in the US as the \"King of Cable\" after his creation of
the"
## [36] "company that became Comcast, attempted to use his 24% holding in"
## [37] "Telewest to force its merger with NTL. | The bond holders of the"
## [38] "heavily indebted companies resisted then, but are unlikely to do so"
```

```
## [39] "this time around. Virgin still has 5.7bn of debt, and a majority of
its"
## [40] "lenders must agree to the takeover by a deadline currently set for
14"
## [41] "February. | Liberty plans to fund its offer by adding nearly 2bn to"
## [42] "Virgin borrowings. The combined Virgin and Liberty Global will owe
a"
## [43] "total of $39bn, more than twice its annual revenues of $17bn, and
more"
## [44] "than its stock market value of $28bn. | Malone, who is now 71, said"
## [45] "yesterday he had no intention to topple his 81-year-old sometime"
## [46] "adversary Rupert Murdoch's UK pay TV business BSkyB, which has
nearly"
## [47] "11 million customers compared with Virgin's 5 million. \"Our
relations"
## [48] "with Sky are going to be very important for us,\" he said. \"We've
had a"
## [49] "long history of cooperation with News Corporation in its various"
## [50] "configurations and we are looking forward to this.\" | Liberty said
it"
## [51] "would continue to invest in Virgin's network, improving its speed
and"
## [52] "capacity to carry data - the top speed offered to retail customers
is"
## [53] "currently 120 megabits a second. But Liberty also hopes to use its"
## [54] "expertise to improve take-up of television and mobile phone"
## [55] "subscriptions at its businesses on the continent. | Malone's
European"
## [56] "broadband and cable companies have 18 million pay-TV subscribers,"
## [57] "almost on a par with Murdoch's 19 million in Britain and Europe."
## [58] "\"Liberty is an 800-pound gorilla and Murdoch is another 800-pound"
## [59] "gorilla,\" said Ovum analyst Adrian Drury. \"The reality is this
will be"
## [60] "a straight bloodbath because the UK is an incredibly competitive"
## [61] "market.\" | The deal is expected to be finalised in June, at which
point"
## [62] "Berkett intends to step down. \"I'm not a very good number two,\"
he"
## [63] "said. | Captions: | 2.6bn | Amount in pounds of historic losses the
new"
## [64] "company will offset against profits - allowing it not to pay"
## [65] "corporation tax for 15 years"
```