# Topic Recommender

## Parameter

Topic cutoff = 15

Num of neighbours = 5

Sim function = topic-based

## Success Rate

```
setwd("/Users/juri/Desktop/RFiles/Results_k15_n5/")
sr_r1 <-  read.csv("SR_Round1", sep="\t", header = F)
sr_r2 <-  read.csv("SR_Round2", sep="\t", header = F)
sr_r3 <-  read.csv("SR_Round3", sep="\t", header = F)
sr_r4 <-  read.csv("SR_Round4", sep="\t", header = F)
sr_r5 <-  read.csv("SR_Round5", sep="\t", header = F)
sr_r6 <-  read.csv("SR_Round6", sep="\t", header = F)
sr_r7 <-  read.csv("SR_Round7", sep="\t", header = F)
sr_r8 <-  read.csv("SR_Round8", sep="\t", header = F)
sr_r9 <-  read.csv("SR_Round9", sep="\t", header = F)
sr_r10 <-  read.csv("SR_Round10", sep="\t", header = F)
sr_tot <- rbind(sr_r1, sr_r2, sr_r3, sr_r4, sr_r5, sr_r6, sr_r7, sr_r8, sr_r9, sr_r10)
sr_tot$V3 <- NULL
sr_tot <- sr_tot %>% group_by(sr_tot$V1) %>% summarise(mean = mean(V2))

sr_tot <- sr_tot %>% rename(k = `sr_tot$V1`)
sr_tot
```

```
## # A tibble: 20 x 2
##        k  mean
##    <int> <dbl>
##  1     1 0.352
##  2     2 0.485
##  3     3 0.556
##  4     4 0.6
##  5     5 0.634
##  6     6 0.658
##  7     7 0.676
##  8     8 0.692
##  9     9 0.707
## 10    10 0.714
## 11    11 0.720
## 12    12 0.718
## 13    13 0.713
## 14    14 0.688
## 15    15 0.659
```

```
## 16    16 0.626
## 17    17 0.574
## 18    18 0.526
## 19    19 0.469
## 20    20 0.412
```

```
sr_tot$mean %>% summary()
```

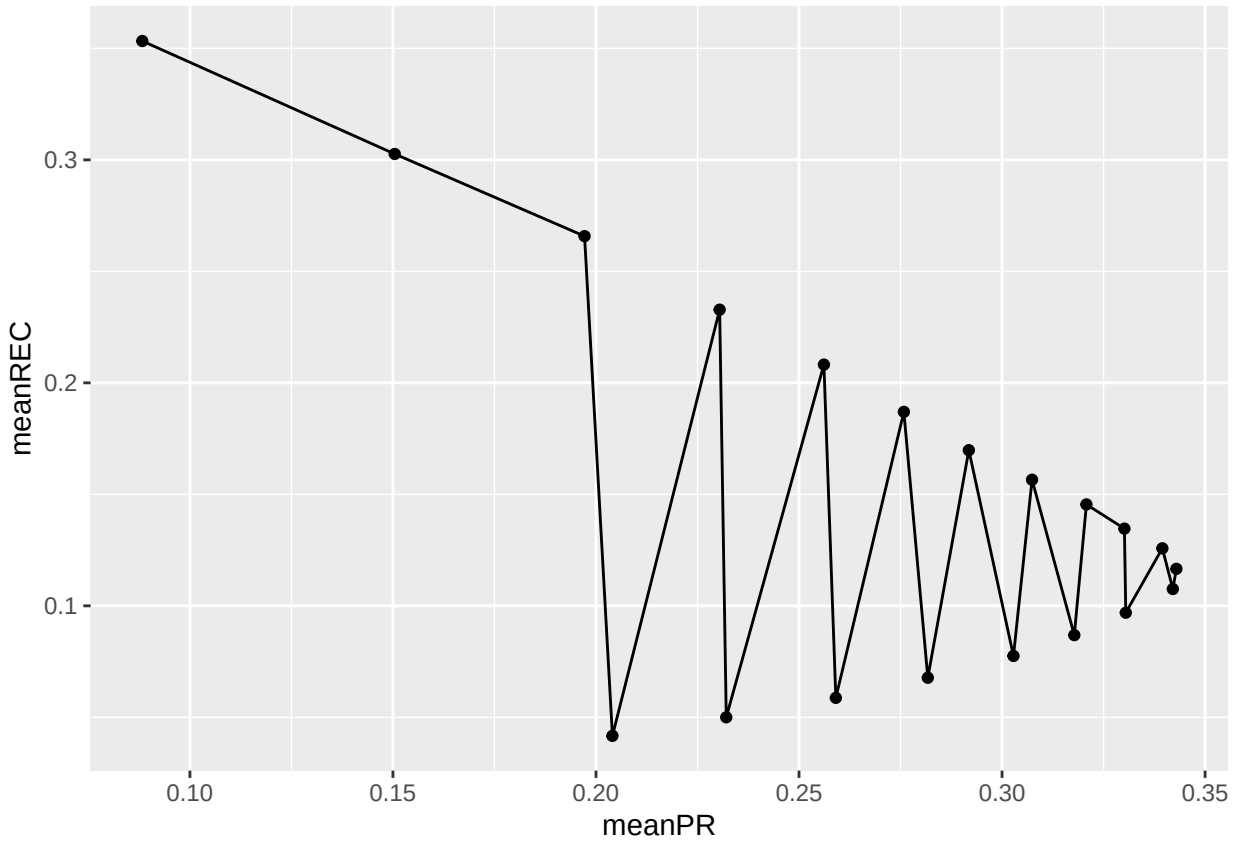
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3520  0.5485  0.6460  0.6090  0.6956  0.7198
```

## Precision and Recall

```
setwd("/Users/juri/Desktop/RFiles/Results_k15_n5/")
pr_r1 <-  read.csv("PRC_Round1", sep="\t", header = F)
pr_r2 <-  read.csv("PRC_Round2", sep="\t", header = F)
pr_r3 <-  read.csv("PRC_Round3", sep="\t", header = F)
pr_r4 <-  read.csv("PRC_Round4", sep="\t", header = F)
pr_r5 <-  read.csv("PRC_Round5", sep="\t", header = F)
pr_r6 <-  read.csv("PRC_Round6", sep="\t", header = F)
pr_r7 <-  read.csv("PRC_Round7", sep="\t", header = F)
pr_r8 <-  read.csv("PRC_Round8", sep="\t", header = F)
pr_r9 <-  read.csv("PRC_Round9", sep="\t", header = F)
pr_r10 <-  read.csv("PRC_Round10", sep="\t", header = F)
pr_tot <- rbind(pr_r1, pr_r2, pr_r3, pr_r4, pr_r5, pr_r6, pr_r7, pr_r8, pr_r9, pr_r10)
pr_tot <- pr_tot %>% group_by(pr_tot$V1) %>% summarise(meanPR = mean(V2), meanREC = mean(V3)) %>% renam
pr_tot
```

```
## # A tibble: 20 x 3
##        k meanPR meanREC
##    <int>  <dbl>   <dbl>
##  1     1 0.0883  0.353
##  2     2 0.150   0.303
##  3     3 0.197   0.266
##  4     4 0.230   0.233
##  5     5 0.256   0.208
##  6     6 0.276   0.187
##  7     7 0.292   0.170
##  8     8 0.307   0.157
##  9     9 0.321   0.145
## 10    10 0.330   0.135
## 11    11 0.339   0.126
## 12    12 0.343   0.117
## 13    13 0.342   0.108
## 14    14 0.330   0.0969
## 15    15 0.318   0.0869
## 16    16 0.303   0.0775
## 17    17 0.282   0.0678
## 18    18 0.259   0.0587
## 19    19 0.232   0.0500
## 20    20 0.204   0.0417
```

```
pr_tot %>% ggplot(aes(x=meanPR, y=meanREC)) + geom_line() + geom_point()
```




**Precision summary**

```
pr_tot$meanPR %>% summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.08828 0.23167 0.28677 0.27005 0.32312 0.34293
```

**Recall summary**

```
pr_tot$meanREC %>% summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.04170 0.08454 0.13021 0.14928 0.19230 0.35330
```