Topic Recommender

Parameter

Topic cutoff = 15

Num of neighbours = 10

Sim function = topic-based

Success Rate

```
setwd("/Users/juri/Desktop/RFiles/Results_k15_n10/")
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 0.380
##  2     2 0.520
##  3     3 0.595
##  4     4 0.648
##  5     5 0.686
##  6     6 0.711
##  7     7 0.733
##  8     8 0.750
##  9     9 0.764
## 10    10 0.774
## 11    11 0.782
## 12    12 0.789
## 13    13 0.796
## 14    14 0.803
## 15    15 0.811
```

```
## 16    16 0.818
## 17    17 0.822
## 18    18 0.825
## 19    19 0.826
## 20    20 0.823
```

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

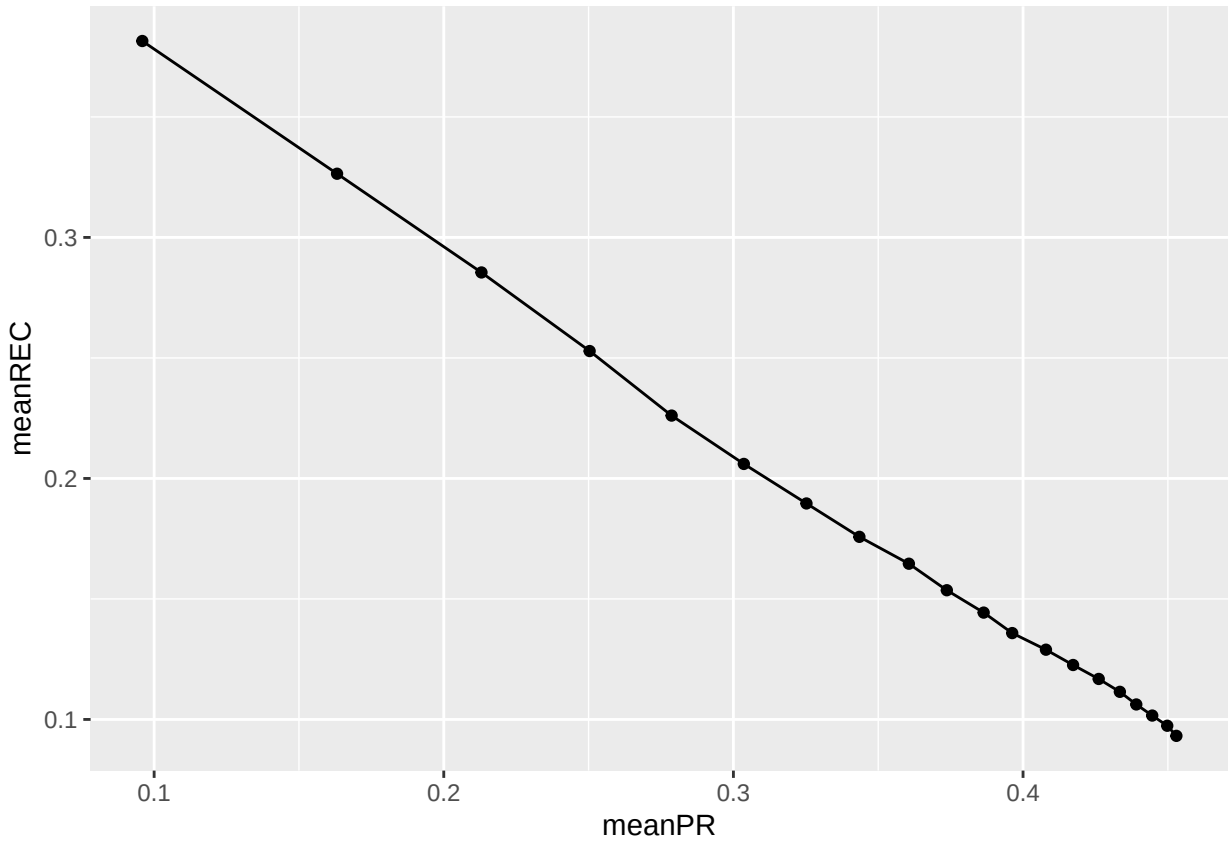
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3802  0.7048  0.7782  0.7329  0.8128  0.8264
```

Precision and Recall

```
setwd("/Users/juri/Desktop/RFiles/Results_k15_n10/")
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 0.0959  0.381
##  2     2 0.163   0.326
##  3     3 0.213   0.285
##  4     4 0.250   0.253
##  5     5 0.279   0.226
##  6     6 0.304   0.206
##  7     7 0.325   0.190
##  8     8 0.344   0.176
##  9     9 0.361   0.165
## 10    10 0.374   0.154
## 11    11 0.386   0.144
## 12    12 0.396   0.136
## 13    13 0.408   0.129
## 14    14 0.417   0.123
## 15    15 0.426   0.117
## 16    16 0.433   0.111
## 17    17 0.439   0.106
## 18    18 0.445   0.102
## 19    19 0.450   0.0974
## 20    20 0.453   0.0932
```

```
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.
## 0.09592 0.29737 0.38005 0.34808 0.42796 0.45294
```

Recall summary

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.09319 0.11546 0.14897 0.17602 0.21104 0.38150
```