Topic Recommender

Parameter

Topic cutoff = 15

Num of neighbours = 25

Sim function = topic-based

Success Rate

```
setwd("/Users/juri/Desktop/RFiles/Results_k15_n25/")
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 0.403
##  2     2 0.540
##  3     3 0.622
##  4     4 0.680
##  5     5 0.723
##  6     6 0.751
##  7     7 0.769
##  8     8 0.788
##  9     9 0.802
## 10    10 0.815
## 11    11 0.829
## 12    12 0.837
## 13    13 0.845
## 14    14 0.854
## 15    15 0.859
```

```
## 16    16 0.865
## 17    17 0.870
## 18    18 0.874
## 19    19 0.880
## 20    20 0.881
```

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

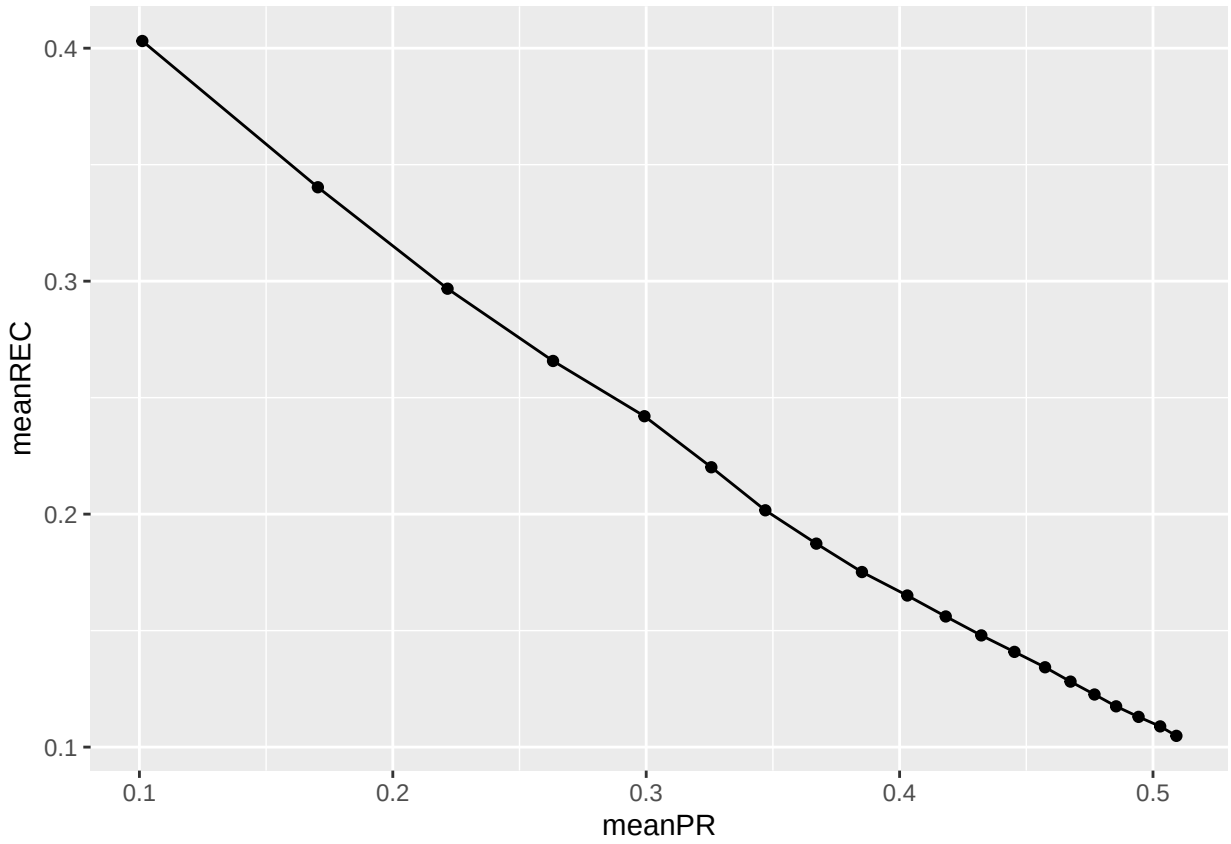
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.4026  0.7437  0.8222  0.7743  0.8605  0.8811
```

Precision and Recall

```
setwd("/Users/juri/Desktop/RFiles/Results_k15_n25/")
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  0.101   0.403
##  2     2  0.170   0.340
##  3     3  0.222   0.297
##  4     4  0.263   0.266
##  5     5  0.299   0.242
##  6     6  0.326   0.220
##  7     7  0.347   0.202
##  8     8  0.367   0.187
##  9     9  0.385   0.175
## 10    10  0.403   0.165
## 11    11  0.418   0.156
## 12    12  0.432   0.148
## 13    13  0.445   0.141
## 14    14  0.457   0.134
## 15    15  0.467   0.128
## 16    16  0.477   0.123
## 17    17  0.485   0.117
## 18    18  0.494   0.113
## 19    19  0.503   0.109
## 20    20  0.509   0.105
```

```
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.
##  0.1011  0.3191  0.4106  0.3787  0.4698  0.5092
```

Recall summary

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1048  0.1267  0.1606  0.1886  0.2256  0.4031
```