# Project 4 - Group 6

Zeyu Gan, Virgile Mison, Galen Simmons, Siyuan Yao, Qingyuan Zhang 4/14/2017

## Setup: assign Knitr root directory and load dependencies

We set the knitr root.dir to the project directory and load/install the necessary packages to run our main.Rmd script.

# Section I: Paper 3

"Name Disambiguation in Author Citations using a K-way Spectral Clustering Method"

As discussed in Section 3.1, authors Han, Zha, and Giles use three citation attributes to design features for name disambiguation. Those attributes are:

- · co-author names
- paper titles

##

<chr>>

• publication venue titles

Together these attributes are called a "citation vector." For a dataset with m features, each citation is representated as an m-dimensional vector given by  $M = (\alpha_1, \ldots, \alpha_m)$ , where  $\alpha_i$  is the weight assigned to feature i. Two types of feature weights assignments are profiled: (i) TFIDF and (ii) normalized TF ("NTF").

We demonstrate how we create the citation vector from our clean data sources in the output library in the chunks below: ### Step 1: Load text file

```
source(file.path(projDir,"lib","feature_extraction.R"))
exFile <- file.path(projDir, "output", "Agupta.csv")</pre>
exText <- readTextFile(exFile)</pre>
## Warning: Missing column names filled in: 'X1' [1]
## Parsed with column specification:
## cols(
##
     X1 = col_integer(),
##
     AuthorID = col_integer(),
##
     PaperNO = col_integer(),
     QuestAuthor = col_character(),
##
##
     Coauthor = col_character(),
     Paper = col_character(),
##
##
     Journal = col_character()
## )
exText %>%
  tbl_df() %>%
  select(QuestAuthor, Coauthor, Paper, Journal) %>%
 head()
## # A tibble: 6 × 4
##
     QuestAuthor
                                                                   Coauthor
```

<chr>>

```
## 1
         A Gupta
                                                                     <NA>
## 2
                                                          A Acero; Y Rui
         A Gupta
## 3
                                                      A Acharya; M Tambe
         A Gupta
## 4
         A Gupta
                                                                A Agarwal
## 5
         A Gupta
                                            A Agrawal; N Chaddha; T Meng
         A Gupta A Balachandran; E Sanocki; G Jancke; J Grudin; J Cadiz
## 6
## # ... with 2 more variables: Paper <chr>, Journal <chr>
```

## Step 2: Prep the co-author and journal terms to be included in the document term matrix

We want the individual co-author names and journal names to be considered as single "terms" in our corpus. Therefore, we collapse the spaces separating the unique letters in someone's name, so that it appears to be a term. For example, "C L Zhang" would become "clzhang". Likewise, with journal names, we combine them into a single string without spaces so that each journal is a unique term in our Document Term Matrix.

```
as tibble(exText) %>%
   mutate(x = str replace all(Coauthor, " ", "")) %>%
   mutate(x = str_replace_all(x, ";", " ")) %>%
   mutate(y = str_replace_all(Journal, " ", "")) %>%
    mutate(term_col = tolower(paste(x, y, Paper))) %>%
    select(term_col) %>%
   head()
## # A tibble: 6 × 1
##
                                                                         term_col
##
                                                                            <chr>
## 1 na parleparallelarchitecturesandlanguageseurope stanford dash multiprocesso
## 2 aacero yrui acmmultimedia automatically extracting highlights for tv baseba
## 3 aacharya mtambe ieeetransparalleldistribsyst implementation of production s
## 4 aagarwal sigmetricsmeasurementandmodelingofcomputersystems memory-reference
## 5 aagrawal nchaddha tmeng icmcsinternationalconferencemultimediacomputingands
## 6 abalachandran esanocki gjancke jgrudin jcadiz cscwconferencecomputersupport
```

Step 3: Run our spectral clustering method on the example dataset

We use the tm package to construct citation vectors in the manner described above. We then use the matching\_matrix and performance\_statistics functions in the lib/evaluation\_metrics.R file to benchmark our results.

```
source(file.path(projDir,"lib","SpectralClustering.R"))
source(file.path(projDir,"lib","evaluation_measures.R"))

num_authors <- length(unique(exText$AuthorID))
author_id <- exText$AuthorID
exDTM <- createCitationMatrix(exText)

exTFIDF <- weightTfIdf(exDTM, normalize = FALSE)
exNTF <- weightTf(exDTM)

exResultsTFIDF <- runSpectralClusteringQR(exTFIDF, num_authors)
exResultsNTF <- runSpectralClusteringQR(exNTF, num_authors)

m0 <- matching_matrix(author_id,exResultsTFIDF)
m1 <- matching_matrix(author_id,exResultsNTF)</pre>
```

```
p0 <- performance_statistics(m0)
p1 <- performance_statistics(m1)</pre>
```

#### Step 4: View results

We examine some results from our sample methods to confirm our methodology

```
ex_df <- data.frame(
    study=rep("Agupta", 2),
    method=c("QR Spectral Clustering - TFIDF", "QR Spectral Clustering - NTF"),
    precision=c(p0$precision, p1$precision),
    recall=c(p0$recall, p1$recall),
    f1=c(p0$f1, p1$f1),
    accuracy=c(p0$accuracy, p1$accuracy),
    mcc=c(p0$mcc, p1$mcc)
    )

kable(ex_df)</pre>
```

study	method	precision	recall	f1	accuracy	mcc
0 1	QR Spectral Clustering - TFIDF QR Spectral Clustering - NTF				$0.7591650 \\ 0.7040427$	

Term Frequency Reverse Document Frequency (TF-IDF) Why TF-IDF? Term frequency counts the number of times a word from a corpus apprears in a particular document. Some words appear frequently in all documents (i.e. they are common locally). Other words appear rarely in the corpus (i.e. they are rare globally). Inverse document frequency weights words that appear rarely more heavily by applying the following transformation to the term frequency (tf) vector:

$$log \cdot \frac{\#ofdocs}{1 + \#ofdocsusingword}$$

#### Distance metrics

Scaled Euclidean Distance:

$$d(x_i, x_q) = \sqrt{a_1(x_i[1] - x_q[1])^2 + \ldots + a_d(x_i[d] - x_q[d])^2}$$

Cosine Similarity:

$$\frac{x_i^T x_q}{\|x_i\| \|x_q\|} = \cos(\theta)$$

Normalizing - normalizing can make dissimilar objects appear more similar. For example a short tweet might appear more similar to a long document.

## Complexity of brute-force search

Given a query point, O(N) distance computations per 1-NN query. By extension,  $O(N \log k)$  distance computations per k-NN query.

#### K-means algorithm

- 1. Initialize cluster centers
- 2. Assign observations to closest cluster center
- 3. Revise cluster centers as mean of assigned observations
- 4. Repeat steps 1 & 2 until convergence

#### Failure modes of k-means

- disparate cluster sizes
- overlapping clusters
- different shaped / oriented clusters

## BELOW IS OLD CODE FROM TEACHER.

This file is currently a template for implementing one of the suggested papers, Han, Zha, & Giles (2005). Due to the nature of the method, we only implement the method on a subset of the data, "AKumar.txt". In your project, you need to work on the whole dataset. You should follow the same structure as in this tutorial, but update it according to the papers you are assigned.

# Step 0: Load the packages, specify directories

```
if (!require("pacman")) install.packages("pacman")

## Loading required package: pacman

pacman::p_load(text2vec, dplyr, qlcMatrix, kernlab, knitr)

#setwd("~/Dropbox/Project4_WhoIsWho/doc")

# here replace it with your own path or manually set it in RStudio
# to where this rmd file is located
```

## Step 1: Load and process the data

For each record in the dataset, there are some information we want to extract and store them in a regular form: canonical author id, coauthors, paper title, publication venue title. You may need to find regular matched in the input string vectors by using regex in R. Here is a tutorial for regular expression in R, which might help you https://rstudio-pubs-static.s3.amazonaws.com/74603 76cd14d5983f47408fdf0b323550b846.html

```
# extract canonical author id befor "_"
AKumar$AuthorID <- sub("_.*","",AKumar$Coauthor)
# extract paper number under same author between "_" and first whitespace
AKumar$PaperNO <- sub(".*_(\\w*)\\s.*", "\\1", AKumar$Coauthor)
# delete "<" in AKumar$Coauthor, you may need to further process the coauthor
# term depending on the method you are using
AKumar$Coauthor <- gsub("<","",sub("^.*?\\s","", AKumar$Coauthor))
# delete "<" in AKumar$Paper
AKumar$Paper <- gsub("<","",AKumar$Paper)
# add PaperID for furthur use, you may want to combine all the nameset files and
# then assign the unique ID for all the citations
AKumar$PaperID <- rownames(AKumar)</pre>
```

# Step 2: Feature design

Following the section 3.1 in the paper, we want to use paper titles to design features for citations. As the notation used in the paper, we want to find a m-dimensional citation vector  $\alpha_i$  for each citation i, i = 1, ..., n. In this dataset, n = 244. We study "TF-IDF" (term frequency-inverse document frequency) as suggested in the paper.

TF-IDF is a numerical statistics that is intended to reflect how important a word is to a document in a collection or corpus. It is often used as a weighting factor in information retrieval, text mining, and user modeling. The TF-IDF value increases proportionally to the number of times a word appears in the document, but is offset by the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general.

```
\begin{aligned} \text{TF}(t) &= \frac{\text{Number of times term } t \text{ appears in a document}}{\text{Total number of terms in the document}} \\ \text{IDF}(t) &= \log \frac{\text{Total number of documents}}{\text{Number of documents with term } t \text{ in it}} \\ \text{TF-IDF}(t) &= \text{TF}(t) \times \text{IDF}(t) \end{aligned}
```

To compute TF-IDF, we first need to construct a document-term matrix (DTM). In other words, the first step is to vectorize text by creating a map from words to a vector space. There are some good packages you could use for text mining (probably you have tried during first project, you don't need to follow my code if you are familiar with other package), e.g. text2vec, tm, tidytext. Here, we are going to use text2vec package. A good tutorial can be found here, https://cran.r-project.org/web/packages/text2vec/vignettes/text-vectorization.html.

Let???s first create a vocabulary-based DTM. Here we collect unique terms from all documents and mark each of them with a unique ID using the create\_vocabulary() function. We use an iterator to create the vocabulary.

## Number of docs: 244

```
## 9 stopwords: a, an, the, in, on, at ...
## ngram_min = 1; ngram_max = 1
## Vocabulary:
##
                terms terms_counts doc_counts
##
     1:
             modular
                                  1
                                              1
     2: interference
                                              1
##
                                  1
##
     3:
            assembly
                                  1
                                              1
##
     4:
           tolerance
                                  1
                                              1
##
     5:
                 form
                                  1
                                              1
##
## 662:
                based
                                 40
                                             38
## 663:
                yeast
                                  1
                                              1
## 664:
              elastic
                                  2
                                              2
## 665:
              schemes
                                  1
                                              1
## 666:
                                  3
                                              3
         information
```

Here, we remove pre-defined stopwords, the words like ???a???, ???the???, ???in???, ???I???, ???you???, ???on???, etc, which do not provide much useful information.

Now that we have a vocabulary list, we can construct a document-term matrix.

```
vectorizer <- vocab_vectorizer(vocab)
dtm_train <- create_dtm(it_train, vectorizer)</pre>
```

Now we have DTM and can check its dimensions.

```
dim(dtm_train)
```

```
## [1] 244 666
```

As you can see, the DTM has 244 rows, equal to the number of citations, and 666, equal to the number of unique terms excluding stopwords.

Then, we want to use DTM to compute TF-IDF transformation on DTM.

```
tfidf <- TfIdf$new()
dtm_train_tfidf <- fit_transform(dtm_train, tfidf)</pre>
```

# Step 3: Clustering

Following suggestion in the paper, we carry out spectral clustering on the Gram matrix of the citation vectors by using R function specc() in *kernlab*. The number of clusters is assumed known as stated in the paper.

```
## Spectral Clustering object of class "specc"
##
## Cluster memberships:
##
## 8 8 8 2 5 11 2 2 2 12 2 11 4 2 1 12 2 14 6 12 14 3 3 7 12 6 2 3 12 12 6 12 1 7 11 11 12 11 12 12 7 1
##
## Gaussian Radial Basis kernel function.
## Hyperparameter: sigma = 0.0474582700433533
```

```
##
Centers:
##
##
     [,2]
       [,3]
         [,4]
##
##
[2,] 0.01112042 0.01112042 0.01112042 0.01429768 0.01429768 0.01429768
##
##
##
##
##
##
   [,7]
     [8,]
      [,9]
        [,10]
          [,11]
[1,] 0.00000000 0.00000000 0.02053 0.02053 0.00000000 0.00000000
##
##
[2,] 0.02502094 0.01668063 0.00000 0.00000 0.01668063 0.00000000
##
##
##
##
##
##
##
  [,13]
     [,14]
       [,15]
         [,16]
[1,] 0.00000000 0.00000000 0.02416789 0.00000000 0.00000000 0.00000000
##
##
##
##
##
##
[14,] 0.00000000 0.02729557 0.02499180 0.03336126 0.00000000 0.03753141
       [,21]
         [,22]
            [,23]
##
  [,19]
     [,20]
##
##
[2,] 0.00000000 0.02175066 0.01309094 0.009098525 0.009098525 0.01429768
##
##
[4,] 0.00000000 0.00000000 0.00000000 0.000000000 0.000000000 0.000000000
##
```

```
##
[14,] 0.03753141 0.03211620 0.03436372 0.000000000 0.000000000 0.00000000
##
  [,25]
    [,26]
      [,27]
        [,28]
         [,29]
           [,30]
##
[2,] 0.01429768 0.01429768 0.01429768 0.01429768 0.0283637 0.0000000
##
##
##
##
##
##
##
  Γ.317
    [.32]
      [,33]
        [.34]
         [.35]
[1,] 0.00000000 0.03695401 0.03695401 0.03417681 0.02053 0.03417681
##
##
##
##
##
##
##
  [,37]
    [,38]
      [,39]
        [,40]
         [,41]
##
           [,42]
##
[2,] 0.00000000 0.00000000 0.01429768 0.0000000 0.000000 0.00000000
##
##
##
[5,] 0.00000000 0.00000000 0.00000000 0.1201005 0.000000 0.09608042
##
[13,] 0.08006702 0.00000000 0.00000000 0.0000000 0.106756 0.00000000
##
  [,43]
    [,44]
      [,45]
        [,46]
         [,47]
           [,48]
```

```
[1,] 0.00000000 0.00000000 0.06159001 0.06159001 0.0000000 0.1159477
##
##
##
##
##
##
##
    [,50]
     [,51]
       [,52]
        [,53]
##
##
##
##
[6,] 0.00000000 0.00000000 0.0000000 0.0000000 0.00000000 0.00000000
##
[10,] 0.00000000 0.00000000 0.0593089 0.0593089 0.10860632 0.08896335
##
  [,55]
##
    [,56]
     [,57]
       [,58]
         [,59]
##
##
##
[3,] 0.00000000 0.06862887 0.06862887 0.00000000 0.00000000 0.00000000
##
##
##
##
[10,] 0.08145474 0.00000000 0.00000000 0.06672251 0.06672251 0.00000000
##
    [,62]
     [,63]
       [,64]
         [,65]
##
##
##
[4,] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##
##
##
##
```

```
[12,] 0.00000000 0.02965445 0.02965445 0.02965445 0.02715158 0.00000000
[,67]
        [,70]
##
    [,68]
      [,69]
          [,71]
            [,72]
##
##
##
##
##
[13,] 0.00000000 0.00000000 0.288608 0.1482164 0.02668901 0.02668901
##
  [,73]
     [,74]
       [,75]
         [,76]
           [,77]
             [,78]
##
##
##
[5,] 0.000000000 0.00000000 0.08006702 0.18218260 0.0000000 0.0000000
##
[12,] 0.000000000 0.00000000 0.00000000 0.02287289 0.0000000 0.0000000
[13,] 0.000000000 0.09774569 0.00000000 0.03087841 0.0000000 0.0000000
##
      [,81]
    [.80]
           [.83]
##
  [,79]
         [,82]
             Γ.847
##
##
##
[5,] 0.02528432 0.02315029 0.02528432 0.02528432 0.1612804 0.02528432
##
##
##
##
       [,87]
  [,85]
    [,86]
         [,88]
           [,89]
             [,90]
##
[1,] 0.01537957 0.01421308 0.01421308 0.01421308 0.00000000 0.0000000
```

```
##
[6,] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##
##
[,91]
     [,92]
       [,93]
         [,94]
           [,95]
##
##
[2,] 0.01251047 0.01668063 0.0000000 0.0000000 0.03054553 0.00000000
##
##
##
##
[8,] 0.00000000 0.00000000 0.0000000 0.0000000 0.00000000 0.00000000
##
[10,] 0.00000000 0.00000000 0.0000000 0.0593089 0.00000000 0.13050393
[13,] 0.00000000 0.00000000 0.1482164 0.0000000 0.00000000 0.00000000
[,97]
    [,98]
       [,99]
         [,100]
           [,101]
##
##
[2,] 0.0000000 0.00000000 0.01000838 0.02061823 0.00000000 0.01112042
##
##
##
##
##
##
[,103]
    [,104]
      [,105]
        [,106]
          [,107]
             [,108]
##
##
##
##
[9,] 0.00000000 0.00000000 0.0000000 0.0727882 0.0727882 0.00000000
##
[10,] 0.00000000 0.00000000 0.1779267 0.0000000 0.0000000 0.00000000
```

```
[,112]
##
    [,110]
      [,111]
##
##
[2,] 0.00000000 0.00000000 0.01429768 0.01429768 0.00000000 0.00000000
##
##
##
  [,115]
    [,116]
      [,117]
        [,118]
          [,119]
##
##
##
##
[14,] 0.00000000 0.03753141 0.06872744 0.06872744 0.06872744 0.107054
##
  [,121]
    [,122]
      [,123]
        [,124]
          [,125]
##
##
[4,] 0.00000000 0.1374549 0.00000000 0.1374549 0.00000000 0.1099639
##
##
[12,] 0.00000000 0.0000000 0.01976963 0.0000000 0.01976963 0.0000000
##
##
  [,127]
    [,128]
     [,129]
        [,130]
          [,131]
[1,] 0.0000000 0.0000000 0.0000000 0.03079501 0.00000000 0.00000000
##
[4,] 0.1099639 0.1099639 0.1099639 0.00000000 0.00000000 0.02729557
##
[6,] 0.0000000 0.0000000 0.0000000 0.00000000 0.00000000 0.00000000
```

```
##
##
  [,133]
   [,134]
     [,135]
       [,136]
         [,137]
           [,138]
##
[2,] 0.0000000 0.0000000 0.0000000 0.02502094 0.02502094 0.00000000
##
##
##
##
##
##
[9,] 0.2802869 0.2802869 0.2802869 0.00000000 0.00000000 0.000000000
##
        [.142]
##
  Γ.1397
    Γ.140]
      [,141]
          Γ.1437
##
[2,] 0.00000000 0.00000000 0.00000000 0.02061823 0.00000000 0.02061823
##
##
##
##
##
[14,] 0.00000000 0.00000000 0.03753141 0.00000000 0.00000000 0.00000000
##
  [,145]
    [,146]
      [,147]
        [,148]
          [,149]
##
            [,150]
##
[2,] 0.00000000 0.04582144 0.02502094 0.0000000 0.00000000 0.01429768
##
##
##
##
[14,] 0.02749097 0.04235993 0.00000000 0.1120422 0.03753141 0.00000000
##
  [,151]
    [,152]
      [,153]
        [,154]
          [,155]
            [,156]
```

```
[2,] 0.0000000 0.00000000 0.02558686 0.009098525 0.009098525 0.009098525
##
##
##
##
##
[8,] 0.0000000 0.00000000 0.00000000 0.000000000 0.000000000 0.000000000
[11,] 0.0000000 0.00000000 0.08099438 0.000000000 0.000000000 0.000000000
##
##
     [,158]
  [,157]
       [,159]
         [,160]
           [,161]
##
[2,] 0.00000000 0.009098525 0.0000000 0.009098525 0.02360336 0.01822374
##
##
##
##
[10,] 0.00000000 0.000000000 0.0762543 0.000000000 0.00000000 0.00000000
##
  [,163]
         [,166]
##
    [,164]
       [,165]
            [,167]
##
[1,] 0.05137632 0.04511339 0.000000000 0.000000000 0.000000000
##
[2,] 0.00000000 0.00000000 0.007148841 0.007148841 0.009098525
##
##
##
##
##
     [,169]
       [,170]
          [,171]
##
##
[2,] 0.007148841 0.007148841 0.007148841 0.008340314 0.01000838
##
##
##
##
##
```

```
##
[,175]
           [,176]
              [,177]
##
   [,173]
      [,174]
                 [,178]
 [1,] 0.00000000 0.000000000 0.00000000 0.01879725 0.00000000 0.00000000
 [2,] 0.01680004 0.008340314 0.01000838 0.01527276 0.00000000 0.00000000
##
##
 ##
 [6,] 0.00000000 0.000000000 0.00000000 0.00000000 0.00000000 0.00000000
##
 [8,] 0.00000000 0.000000000 0.00000000 0.04893897 0.00000000
##
 ##
   [,179]
      [,180]
        [,181]
           [,182]
              [,183]
                 [,184]
 ##
 [2,] 0.007698752 0.0000000 0.00000000 0.007698752 0.00000000 0.007698752
##
 ##
 ##
 [12,] 0.000000000 0.0000000 0.09359351 0.000000000 0.00000000 0.000000000
##
      Γ.1867
         [,187]
           [,188]
              Γ.1897
##
   [,185]
                 [.190]
##
 [2,] 0.01309094 0.005887281 0.00000000 0.01000838 0.00000000 0.01000838
##
 [4,] 0.02290915 0.000000000 0.02502094 0.00000000 0.00000000 0.00000000
##
 ##
 ##
 ##
 ##
##
   [,191]
      [,192]
        [,193]
           [,194]
              [,195]
                 [,196]
##
 [1,] 0.00000000 0.00000000 0.01216235 0.00000000 0.000000000 0.00000000
 [2,] 0.00000000 0.00000000 0.01712864 0.03665463 0.008340314 0.00000000
##
```

```
##
##
##
[,198]
        [,199]
          [,200]
   [,197]
            [,201]
##
##
[1,] 0.000000000 0.00000000 0.000000000 0.01661423 0.00000000 0.00000000
[2,] 0.008340314 0.01832732 0.008340314 0.00000000 0.01000838 0.01309094
##
##
##
##
##
[14.] 0.000000000 0.00000000 0.000000000 0.07608563 0.00000000 0.00000000
  [,203]
     [,204]
       [,205]
          [,206]
            [,207]
##
##
[2,] 0.01000838 0.007148841 0.02001675 0.00000000 0.008340314 0.007148841
##
##
[4,] 0.00000000 0.000000000 0.00000000 0.00000000 0.000000000 0.000000000
##
##
##
##
[11,] 0.00000000 0.000000000 0.00000000 0.04448168 0.000000000 0.000000000
[,209]
     [,210]
       [,211]
          [,212]
            [,213]
##
##
[2,] 0.00000000 0.01000838 0.01712864 0.00000000 0.008340314 0.008340314
##
##
[9,] 0.00000000 0.00000000 0.05338797 0.00000000 0.000000000 0.000000000
##
[11,] 0.00000000 0.00000000 0.00000000 0.02859536 0.000000000 0.000000000
```

```
[,218]
##
      [,216]
        [,217]
 ##
##
 [2,] 0.007148841 0.00000000 0.000000000 0.00000000 0.01309094 0.00000000
 ##
 ##
 [9,] 0.000000000 0.00000000 0.000000000 0.04367292 0.00000000 0.00000000
[12,] 0.000000000 0.01617516 0.008896335 0.00000000 0.00000000 0.08145474
##
   [,221]
      [,222]
        [,223]
           [,224]
             [,225]
 ##
 [2,] 0.009163658 0.00000000 0.01668063 0.00000000 0.0000000 0.02867281
##
 [4,] 0.000000000 0.00000000 0.00000000 0.02729557 0.0000000 0.02105346
 ##
 ##
 ##
   [,227]
      [,228]
        [,229]
           [,230]
              [,231]
 ##
 ##
 [4,] 0.000000000 0.00000000 0.03336126 0.00000000 0.00000000 0.00000000
 ##
 ##
 [14,] 0.000000000 0.00000000 0.00000000 0.03753141 0.00000000 0.00000000
##
  [,233]
     [,234]
        [,235]
          [,236]
             [,237]
##
 [1,] 0.0153975 0.00000000 0.000000000 0.0153975 0.00000000 0.01421308
 [2,] 0.0000000 0.00000000 0.007148841 0.0000000 0.00000000 0.00000000
 ##
 [4,] 0.0000000 0.02729557 0.000000000 0.0000000 0.02729557 0.00000000
##
 ##
```

```
##
##
  [,239]
     [,240]
        [,241]
          [,242]
             [,243]
               [,244]
[1,] 0.00000000 0.000000000 0.00000000 0.01421308 0.04357544 0.00000000
##
[2,] 0.01429768 0.009163658 0.00000000 0.00000000 0.00000000 0.01668063
##
##
[4,] 0.00000000 0.024991795 0.03336126 0.00000000 0.00000000 0.00000000
##
##
##
##
##
        [,247]
          [,248] [,249]
     [,246]
   [.245]
[1,] 0.00000000 0.018797248 0.00000000 0.00000000 0.02053 0.00000000
##
##
##
##
##
##
[8,] \quad 0.00000000 \quad 0.000000000 \quad 0.04448168 \quad 0.00000000 \quad 0.00000 \quad 0.00000000
[9,] 0.04367292 0.000000000 0.00000000 0.03970265 0.00000 0.00000000
##
  [,251]
     [,252]
       [,253]
          [,254]
            [,255]
##
               [,256]
##
[2,] 0.0000000 0.02001675 0.01668063 0.00000000 0.01619888 0.00000000
##
##
##
##
[9,] 0.0727034 0.00000000 0.00000000 0.03970265 0.06426001 0.00000000
##
  [,257]
     [,258]
        [,259]
          [,260]
             [,261]
               [,262]
```

```
[2,] 0.01372487 0.009515912 0.01832732 0.01000838 0.00000000 0.00000000
##
 ##
##
 ##
 ##
 [8,] 0.00000000 0.000000000 0.00000000 0.00000000 0.00000000 0.00000000
 ##
##
         [,265]
            [,266]
   [,263]
      [,264]
               [,267]
##
 ##
 ##
 ##
  \hbox{ \tt [6,] } \hbox{ \tt 0.00000000 } \hbox{ \tt 0.00000000 } \hbox{ \tt 0.00000000 } \hbox{ \tt 0.07330927 } \hbox{ \tt 0.00000000 } \hbox{ \tt 0.00000000} 
 ##
 [11,] 0.00000000 0.00000000 0.04581829 0.00000000 0.00000000 0.00000000
[12,] 0.03846474 0.01976963 0.02715158 0.00000000 0.04072737 0.00000000
##
##
   [,269]
      [,270]
         [,271]
           [,272]
               [,273]
                  [,274]
##
 [1,] 0.00000000 0.0000000 0.00000000 0.02035939 0.000000000 0.00000000
##
 [2,] 0.00000000 0.0000000 0.00000000 0.01543920 0.009098525 0.00000000
 ##
 ##
##
 ##
 [11,] 0.05301722 0.0000000 0.05672741 0.00000000 0.000000000 0.04072737
[14,] 0.00000000 0.0000000 0.00000000 0.07251747 0.000000000 0.00000000
##
      [,276]
         [,277]
            [,278]
               [,279]
 ##
##
 [2,] 0.02061823 0.00000000 0.00000000 0.01251047 0.00000000 0.008340314
 ##
 ##
 ##
##
 ##
```

```
[11,] 0.00000000 0.00000000 0.04448168 0.00000000 0.00000000 0.000000000
[,282]
        [,283]
           [,284]
              [,285]
##
   [,281]
 [1.] 0.02819587 0.00000000 0.02053 0.00000000 0.00000000 0.00000000
 [2,] 0.00000000 0.02197924 0.00000 0.00000000 0.01917176 0.00000000
##
##
 [4,] 0.00000000 0.00000000 0.00000 0.03753141 0.00000000 0.00000000
##
 ##
 ##
 [10,] 0.08145474 0.00000000 0.00000 0.00000000 0.04574579 0.00000000
##
   [,287]
      [,288]
        [,289]
           [,290]
              [,291]
                 [,292]
 ##
 ##
 [4,] 0.03753141 0.0000000 0.00000 0.000000000 0.01815104 0.00000000
##
 ##
 [7,] 0.00000000 0.0000000 0.23068 0.000000000 0.25665882 0.00000000
 [12,] 0.00000000 0.0000000 0.00000 0.008896335 0.00000000 0.00000000
##
      [,294]
         [,295]
            [,296]
##
   [.293]
               [,297]
                  [,298]
##
 [2,] 0.0000000 0.00000000 0.0000000 0.02443642 0.01325431 0.00654547
##
 ##
 ##
 ##
 [11,] 0.1446328 0.05004189 0.1243866 0.00000000 0.00000000 0.00000000
##
[,300]
         [,301]
               [,303]
##
   [,299]
            [,302]
                  [,304]
##
 [2,] 0.00000000 0.00000000 0.00654547 0.00000000 0.00000000 0.00000000
```

```
[4,] 0.00000000 0.00000000 0.02499180 0.00000000 0.00000000 0.00000000
##
[6,] 0.04575258 0.00000000 0.00000000 0.04575258 0.00000000 0.00000000
##
##
[,306]
       [,307]
         [,308]
            [,309]
##
  [,305]
##
##
##
[4,] 0.00000000 0.0000000 0.00000000 0.02729557 0.00000000 0.00000000
##
[6,] 0.00000000 0.1761803 0.06479551 0.00000000 0.00000000 0.00000000
[7,] 0.00000000 0.0000000 0.05981124 0.00000000 0.00000000 0.00000000
##
##
[,311]
     [,312]
       [,313]
          [,314]
            [,315]
               [,316]
##
##
[2,] 0.00000000 0.0000000 0.007148841 0.00000000 0.009098525 0.00000000
##
##
##
##
[8,] 0.08145474 0.0000000 0.000000000 0.08145474 0.000000000 0.00000000
##
[,317]
     [,318]
       [,319]
          [,320]
            [,321]
##
##
[2,] 0.0000000 0.006749532 0.00000000 0.01251047 0.000000 0.01251047
##
##
[6,] 0.0000000 0.000000000 0.000000000 0.00000000 0.0000000 0.00000000
##
```

```
[14,] 0.0000000 0.000000000 0.06109105 0.00000000 0.000000 0.00000000
       [,324]
          [,325]
              [,326]
##
 [1,] 0.00000000 0.029931 0.01216235 0.00000000 0.0000000 0.00000000
##
##
 [2,] 0.02502094 0.000000 0.01682277 0.00000000 0.0000000 0.00000000
 [3,] 0.00000000 0.000000 0.00000000 0.04902062 0.0000000 0.0000000
##
 ##
##
 ##
 ##
##
    [,329]
        [,330]
           [,331]
              [,332]
                  [,333]
 ##
 [2,] 0.008577335 0.00654547 0.00000000 0.0000000 0.00000000 0.007136934
##
 [3,] 0.000000000 0.00000000 0.02936338 0.0000000 0.00000000 0.000000000
 ##
 ##
 [9,] 0.000000000 0.03998687 0.12457194 0.1332896 0.00000000 0.000000000
 [11,] 0.000000000 0.00000000 0.00000000 0.15561737 0.00000000
 [12,] 0.027447472 0.00000000 0.00000000 0.0000000 0.02646721 0.016917176
 ##
 ##
    [,335]
       [,336]
           [,337]
              [,338]
                 [,339]
 ##
 ##
 ##
 [6,] 0.00000000 0.0000000 0.05864741 0.0000000 0.00000000 0.000000000
 ##
 [10.] 0.00000000 0.0593089 0.00000000 0.0762543 0.00000000 0.00000000
 ##
##
    [,341]
        [,342]
           [,343]
               [,344]
                  [,345]
 [1,] 0.000000000 0.00000000 0.03492948 0.00000000 0.00000000 0.00000000
 [3,] 0.000000000 0.00000000 0.07339468 0.00000000 0.00000000 0.00000000
##
 ##
 [6,] 0.000000000 0.04891847 0.01614518 0.07099074 0.00000000 0.00000000
```

```
##
[11,] 0.000000000 0.04468515 0.01707663 0.04119998 0.00000000 0.00000000
[12,] 0.000000000 0.22677148 0.10301095 0.06638948 0.00000000 0.00000000
##
   Γ.347]
     [,348]
       [,349]
         [,350]
            [,351]
               Γ.3521
[1,] 0.09719326 0.0000000 0.02053 0.00000000 0.00000000 0.02416789
##
 [2,] 0.00000000 0.0000000 0.00000 0.01223474 0.00000000 0.00000000
##
##
##
##
##
[8,] 0.00000000 0.0000000 0.00000 0.07232093 0.06839526 0.00000000
##
##
        [,355]
          [,356]
     [.354]
   [.353]
            [.357]
               [.358]
##
##
[4,] 0.00000000 0.02729557 0.02729557 0.000000 0.0000000 0.00000000
##
##
##
##
[8,] 0.00000000 0.00000000 0.00000000 0.077382 0.0000000 0.00000000
##
##
   [,359]
     [,360]
        [,361]
           [,362]
             [,363]
##
[2,] 0.007698752 0.00000000 0.00000000 0.01429768 0.02063428 0.02198419
##
##
##
[6,] 0.000000000 0.00000000 0.05864741 0.00000000 0.00000000 0.00000000
##
##
   [,365]
     [,366]
        [,367]
           [,368]
             [,369]
                [,370]
```

```
[2,] 0.005887281 0.00000000 0.00000000 0.01561896 0.00000000 0.00000000
##
 [3,] 0.000000000 0.00000000 0.08976645 0.00000000 0.00000000 0.00000000
 ##
##
 [5,] 0.000000000 0.00000000 0.00000000 0.01869083 0.00000000 0.00000000
##
 ##
 [9.] 0.000000000 0.00000000 0.00000000 0.04035521 0.00000000 0.00000000
 [12,] 0.000000000 0.01482723 0.00000000 0.03945842 0.01617516 0.00000000
 ##
##
           [,373]
              [,374]
    [,371]
        [,372]
                  [,375] [,376]
##
 [2,] 0.007698752 0.01668063 0.005887281 0.0000000 0.0000000 0.000000
##
 ##
 ##
 ##
 [11.] 0.000000000 0.00000000 0.000000000 0.2104935 0.0000000 0.00000
 ##
    [,377]
##
       [,378]
          [,379]
              [,380]
                 [,381]
##
 ##
 [2,] 0.00000000 0.02854774 0.0000000 0.00000000 0.00000000 0.005887281
 ##
 [5,] 0.00000000 0.00000000 0.4601626 0.00000000 0.08006702 0.000000000
 ##
##
 ##
 [9,] 0.00000000 0.04671448 0.0000000 0.05459115 0.00000000 0.000000000
 [13.] 0.00000000 0.00000000 0.1054292 0.00000000 0.00000000 0.000000000
 ##
       [,384]
           [,385]
              [,386]
                  [,387]
 [1,] 0.02190850 0.00000000 0.00000000 0.01879725 0.00000000 0.00000000
##
##
 [2,] 0.02525998 0.00000000 0.00654547 0.00000000 0.00000000 0.00000000
 ##
 [4,] 0.00000000 0.00000000 0.02290915 0.00000000 0.00000000 0.00000000
##
 ##
##
 ##
```

```
[12,] 0.00000000 0.00000000 0.00000000 0.01629095 0.00000000 0.02287289
##
[,389]
       [,390]
        [,391]
            [,392]
               [,393]
##
 [1,] 0.00000000 0.00000000 0.02053 0.000000000 0.0000000 0.0000000
 [2,] 0.02001675 0.01000838 0.00000 0.009098525 0.0000000 0.0000000
##
##
 ##
 [6,] 0.00000000 0.00000000 0.00000 0.000000000 0.0000000 0.0000000
##
 ##
 ##
    [,395]
       [,396]
          [,397]
             [,398]
                 [,399]
                    [,400]
##
 [2,] 0.000000000 0.009098525 0.0000000 0.01112042 0.005887281 0.000000000
##
 [4,] 0.000000000 0.000000000 0.1374549 0.00000000 0.000000000 0.000000000
##
 ##
 ##
       [,402]
          [,403]
             [,404]
                [.405]
##
   [,401]
##
 ##
 [3,] 0.00000000 0.00000000 0.00000000 0.04488322 0.00000000 0.0000000
##
 [4,] 0.00000000 0.00000000 0.00000000 0.03436372 0.00000000 0.0000000
##
 [5,] 0.00000000 0.00000000 0.08221748 0.00000000 0.00000000 0.00000000
 ##
 ##
##
 [12,] 0.02283819 0.00000000 0.03806365 0.00000000 0.01482723 0.0000000
##
[,409]
                [,411]
##
    [,407]
       [,408]
             [,410]
                    [,412]
##
 [2,] 0.004631159 0.00000000 0.01018184 0.0000000 0.00000000 0.007911404
##
 [3,] 0.047634778 0.00000000 0.00000000 0.0000000 0.04488322 0.000000000
```

```
[4,] 0.011367390 0.00000000 0.02290915 0.0000000 0.00000000 0.037219561
 ##
 ##
##
 ##
 [11,] 0.027786954 0.00000000 0.00000000 0.1649458 0.00000000 0.000000000
[12.] 0.031697710 0.02961991 0.00000000 0.0000000 0.02036368 0.00000000
[,414]
         [,415]
            [,416]
               [,417]
##
   [,413]
##
 ##
 ##
##
 ##
 [7,] 0.00000000 0.00000000 0.000000 0.12648843 0.06767009 0.00000000
##
 [8,] 0.00000000 0.00000000 0.000000 0.05709547 0.00000000 0.07612729
##
 [11,] 0.00000000 0.00000000 0.132364 0.00000000 0.00000000 0.00000000
[,419]
      [,420]
          [,421]
             [,422]
                [,423]
##
 ##
 [2,] 0.00000000 0.008340314 0.009098525 0.01429768 0.01112042 0.0000000
##
 ##
 ##
 ##
 ##
 ##
[,425]
      [,426]
         [,427]
            [,428]
               [,429]
##
 [1,] 0.0000000 0.00000000 0.00000000 0.01756784 0.000000000 0.0000000
##
 [2,] 0.0000000 0.00000000 0.01602059 0.00000000 0.008340314 0.0000000
##
 [4,] 0.0000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##
##
 ##
[11,] 0.0000000 0.00000000 0.02190850 0.00000000 0.000000000 0.00000000
## [12,] 0.0325819 0.04671448 0.04109820 0.00000000 0.00000000 0.0254181
```

```
[13,] 0.0000000 0.00000000 0.04556969 0.00000000 0.00000000 0.00000000
 [14,] 0.0000000 0.000000000 0.000000000 0.06423240 0.000000000 0.00000000
##
        [,432]
            [,433]
               [,434]
 [1,] 0.029841752 0.00000000 0.00000000 0.05270351 0.000000000 0.00000000
##
##
 [2,] 0.008174029 0.02175066 0.02850916 0.00000000 0.005887281 0.00000000
 ##
 ##
 [7,] 0.000000000 0.00000000 0.00000000 0.12648843 0.000000000 0.00000000
 ##
    [,437]
        [,438]
            [,439]
               [,440]
                   [,441]
 ##
 [2,] 0.007698752 0.00000000 0.00000000 0.02670099 0.00000000 0.008564321
##
 [4,] 0.000000000 0.00000000 0.05553732 0.00000000 0.00000000 0.000000000
 [5,] 0.000000000 0.00000000 0.00000000 0.05696211 0.00000000 0.000000000
##
 [6,] 0.000000000 0.00000000 0.00000000 0.08699668 0.00000000 0.000000000
 ##
 [9,] 0.000000000 0.04671448 0.00000000 0.00000000 0.00000000 0.067948328
 ##
        [,444]
            [,445]
                [,446]
 ##
 [2,] 0.01751458 0.008340314 0.007148841 0.00000000 0.008099438 0.0000000
 ##
 [5,] 0.00000000 0.000000000 0.000000000 0.00000000 0.00000000 0.00000000
##
 ##
 [12,] 0.00000000 0.000000000 0.000000000 0.01976963 0.028798002 0.0000000
 ##
        [,450]
               [,452]
##
    [,449]
           [,451]
                   [,453]
 [2,] 0.01822374 0.00000000 0.00000000 0.007363126 0.00000000 0.00000000
 ##
 [4,] 0.00000000 0.00000000 0.00000000 0.000000000 0.00000000 0.00000000
 ##
```

```
[7,] 0.00000000 0.00000000 0.06767009 0.000000000 0.07027135 0.00000000
 [8,] 0.00000000 0.00000000 0.00000000 0.128562510 0.00000000 0.00000000
##
 ##
     [,455]
        [,456]
            [,457]
                [.458]
                    [,459]
                        Γ.460]
 ##
 [2,] 0.005887281 0.01112042 0.00000000 0.005887281 0.02001675 0.04057981
 ##
##
 ##
 [6,] 0.000000000 0.00000000 0.08509097 0.000000000 0.00000000 0.02792589
##
##
 [7,] 0.000000000 0.00000000 0.18264949 0.000000000 0.00000000 0.05012340
 [8,] 0.000000000 0.00000000 0.14772739 0.000000000 0.00000000 0.05430035
##
 ##
    Γ.461]
        [.462]
            [.463]
               Γ.464]
                   [.465]
 [1,] 0.00000000 0.0000000 0.031926165 0.0000000 0.00000000 0.000000000
##
 [2.] 0.00000000 0.0000000 0.008564321 0.0000000 0.00000000 0.026652103
 [3,] 0.00000000 0.1719064 0.000000000 0.0000000 0.03431444 0.148132233
##
 ##
 ##
 ##
 [7,] 0.00000000 0.0000000 0.000000000 0.3549537 0.00000000 0.041130355
##
 ##
        [,468]
            [,469]
                [,470]
                   [,471]
##
    [,467]
                       [,472]
 [1,] 0.00000000 0.021908504 0.00000000 0.0000000 0.02819587 0.00000000
##
 [2,] 0.01570913 0.007120264 0.00000000 0.0000000 0.00000000 0.01834007
 [3,] 0.00000000 0.000000000 0.00000000 0.2673732 0.00000000 0.00000000
##
 ##
##
 ##
 [12,] 0.00000000 0.000000000 0.02224084 0.0000000 0.00000000 0.00000000
 [14,] 0.00000000 0.066887328 0.00000000 0.0000000 0.06872744 0.06688733
##
    [,473]
        [,474]
           [,475]
               [,476]
                   [,477]
                       [,478]
```

```
##
 ##
##
 [6,] 0.00000000 0.00000000 0.1421574 0.00000000 0.05331583 0.00000000
##
 ##
 [8,] 0.00000000 0.00000000 0.0000000 0.00000000 0.00000000 0.00000000
 [12,] 0.00000000 0.00000000 0.0000000 0.02965445 0.00000000 0.00000000
 ##
##
                [,481]
           [,480]
                     [,482]
                          [,483]
##
 [1,] 0.01986713 0.009933567 0.00000000 0.00000000 0.01517632 0.00000000
 [2,] 0.00000000 0.020397677 0.01956434 0.007148841 0.01268306 0.00000000
##
 [3,] 0.00000000 0.000000000 0.27264552 0.000000000 0.00000000 0.00000000
 ##
 [5,] 0.00000000 0.000000000 0.01139455 0.000000000 0.00000000 0.00000000
##
  \hbox{ \tt [6,] 0.23385423 0.000000000 0.04198718 0.000000000 0.00000000 0.000000000} \\
 ##
 [9,] 0.00000000 0.000000000 0.03578452 0.000000000 0.00000000 0.00000000
 [10,] 0.00000000 0.000000000 0.04009192 0.000000000 0.04384269 0.00000000
 [11,] 0.00000000 0.096852280 0.00000000 0.000000000 0.00000000 0.05672741
 [12,] 0.01147879 0.035219011 0.03462484 0.000000000 0.00000000 0.00000000
 [14,] 0.00000000 0.048426140 0.00000000 0.000000000 0.07213493 0.00000000
##
     [,485]
##
          [,486]
               [,487]
                    [,488]
                          [,489]
                               [,490]
##
 ##
 [2,] 0.00000000 0.00000000 0.01018184 0.01388865 0.009098525 0.009098525
 [3,] 0.03162523 0.00000000 0.03927282 0.00000000 0.000000000 0.000000000
 ##
 [6,] 0.00000000 0.00000000 0.00000000 0.19399669 0.000000000 0.000000000
##
 ##
 ##
         [,492]
              [,493]
                   [,494]
                        [,495]
 ##
##
 [3,] 0.00000 0.00000000 0.06077927 0.00000000 0.0000000 0.00000000
##
 [4,] 0.00000 0.01718721 0.01450414 0.00000000 0.0000000 0.00000000
 ##
 [6,] 0.00000 0.00000000 0.04254549 0.00000000 0.0000000 0.00000000
##
##
 [8,] 0.00000 0.00000000 0.00000000 0.08700262 0.0000000 0.00000000
##
 [9,] 0.00000 0.00000000 0.00000000 0.05338797 0.0000000 0.04367292
```

```
[12,] 0.00000 0.02444403 0.13394973 0.00000000 0.0254181 0.00000000
 [,497]
             [,498]
                    [,499]
                         [,500]
                               [,501]
##
  [2,] 0.00000000 0.00000000 0.009515912 0.0000000 0.0000000 0.01018184
##
  [4,] 0.00000000 0.00000000 0.044768039 0.0000000 0.0000000 0.00000000
  ##
  [11,] 0.05672741 0.00000000 0.000000000 0.1649458 0.0000000 0.00000000
 [12,] 0.00000000 0.02537576 0.000000000 0.0000000 0.0254181 0.00000000
 ##
        [,503]
              [,504]
                    [,505]
                          [,506]
                                [,507]
                                      [,508]
  [1,] 0.000000000 0.00000000 0.01476527 0.00000000 0.01879725 0.00000000
##
  [2,] 0.005887281 0.01112042 0.05620066 0.01018184 0.00000000 0.00000000
##
  [3,] 0.000000000 0.00000000 0.01980421 0.00000000 0.03141826 0.09256501
  [4,] 0.000000000 0.00000000 0.05308940 0.02290915 0.00000000 0.00000000
##
  [5,] 0.000000000 0.00000000 0.07278045 0.00000000 0.00000000 0.00000000
  [6,] 0.000000000 0.00000000 0.01848392 0.00000000 0.00000000 0.00000000
##
  [8,] 0.000000000 0.00000000 0.07380734 0.00000000 0.00000000 0.00000000
  [10,] 0.000000000 0.00000000 0.01925409 0.00000000 0.00000000 0.00000000
 [11,] 0.000000000 0.00000000 0.10414134 0.00000000 0.00000000 0.00000000
 [12,] 0.000000000 0.00000000 0.05678984 0.00000000 0.00000000 0.03999723
 [13,] 0.000000000 0.00000000 0.01540327 0.00000000 0.00000000 0.00000000
 [14,] 0.000000000 0.00000000 0.04583786 0.00000000 0.00000000 0.00000000
##
                   [,511]
                         [,512]
##
       [,509]
             [,510]
                               ſ.513l
##
  [1,] 0.00000000 0.04229381 0.01253085 0.00000000 0.0000000 0.00000000
  [2,] 0.00000000 0.00000000 0.02332447 0.00000000 0.0000000 0.01251047
##
  [4,] 0.00000000 0.00000000 0.01527197 0.04790094 0.3002513 0.00000000
##
  [6,] 0.00000000 0.00000000 0.03554205 0.00000000 0.0000000 0.00000000
##
  [7,] 0.00000000 0.00000000 0.09523446 0.00000000 0.0000000 0.00000000
  [8,] 0.00000000 0.00000000 0.05158533 0.00000000 0.0000000 0.00000000
##
  [,517]
                         [,518]
                              [,519]
##
       [,515]
             [,516]
                                     [,520]
##
  [1,] 0.04229381 0.01879725 0.02190850 0.00000000 0.000000 0.00000000
  [2,] 0.00000000 0.00000000 0.01649658 0.00000000 0.000000 0.00000000
  [3,] 0.00000000 0.00000000 0.02441233 0.00000000 0.000000 0.00000000
```

```
##
 [6,] 0.00000000 0.00000000 0.00000000 0.00000000 0.0000000 0.00000000
 ##
 [8,] 0.00000000 0.06109105 0.03164562 0.08145474 0.000000 0.000000000
##
 [9,] 0.00000000 0.00000000 0.04438606 0.00000000 0.000000 0.00000000
[,521]
        [,522]
           [,523]
               [,524]
                  [,525]
##
##
 [1,] 0.000000000 0.077185346 0.00000000 0.00000000 0.0153975 0.00000000
 [2,] 0.008577335 0.005085903 0.02181823 0.00000000 0.0000000 0.00000000
##
 [3,] 0.000000000 0.000000000 0.00000000 0.15702842 0.0000000 0.00000000
##
 ##
 [6,] 0.000000000 0.000000000 0.00000000 0.05918763 0.0000000 0.08639401
 ##
 [8,] 0.000000000 0.000000000 0.00000000 0.00000000 0.0000000 0.00000000
##
 [,527]
       [,528]
           [,529]
              [,530]
                  [,531]
                      [,532]
##
 ##
 [2,] 0.005887281 0.01527752 0.00000000 0.02699813 0.008220505 0.01429768
##
 [3,] 0.000000000 0.00000000 0.06862887 0.00000000 0.00000000 0.00000000
##
 [4,] 0.000000000 0.03593689 0.00000000 0.03037289 0.000000000 0.00000000
##
 [6,] 0.000000000 0.04399925 0.00000000 0.03702600 0.000000000 0.00000000
##
 ##
 ##
[12,] 0.000000000 0.02444403 0.00000000 0.00000000 0.006576404 0.00000000
 [,533]
       [,534]
          [,535]
              [,536]
                  [,537]
##
                     [,538]
 ##
 [2,] 0.0222546 0.00000000 0.01251047 0.00000000 0.000000000 0.008564321
##
 ##
 ##
```

```
[14,] 0.0000000 0.00000000 0.00000000 0.08992537 0.039886394 0.000000000
##
      [,540]
         [,541]
            [,542]
 ##
##
 ##
 ##
 [11,] 0.00000000 0.05236376 0.00000000 0.06672251 0.05236376 0.05719073
[12,] 0.00000000 0.00000000 0.02965445 0.00000000 0.00000000 0.00000000
##
   [,545]
      [,546]
         [,547]
            [,548]
                [,549]
 ##
 [2,] 0.0000000 0.009163658 0.01000838 0.00000000 0.060964171 0.00000000
##
 ##
 ##
 [8,] 0.1359416 0.000000000 0.00000000 0.00000000 0.036205506 0.00000000
 [14,] 0.0000000 0.000000000 0.00000000 0.06872744 0.000000000 0.00000000
##
       [,552]
          [,553]
             [,554]
                [,555]
 [1,] 0.00000000 0.000000000 0.00000000 0.01301348 0.00000000 0.01216235
##
 ##
 ##
 [7,] 0.00000000 0.000000000 0.09667156 0.00000000 0.19545013 0.00000000
##
 [8,] 0.00000000 0.000000000 0.00000000 0.03665463 0.03430934 0.08564321
 ##
      [,558]
         [,559]
            [,560]
               [,561]
##
   [,557]
 [1,] 0.00000000 0.00000000 0.00000000 0.1990569 0.00000000 0.000000000
 [2,] 0.00000000 0.01000838 0.01000838 0.0000000 0.03374766 0.008340314
 ##
 [4,] 0.07854564 0.00000000 0.00000000 0.0000000 0.06942376 0.00000000
 ##
```

```
[7,] 0.00000000 0.00000000 0.00000000 0.2107257 0.00000000 0.000000000
##
##
  [,563]
     [,564]
       [,565]
         [,566]
           [,567]
              [,568]
[1,] 0.04229381 0.00000000 0.00000000 0.02053 0.00000000 0.00000000
##
[2,] 0.00000000 0.01112042 0.02749097 0.00000 0.02001675 0.00000000
##
##
##
##
##
##
     [.570]
  [.569]
       [,571]
          [.572]
            [.573]
##
[2,] 0.00000000 0.01112042 0.00000000 0.009515912 0.00000000 0.000000
##
[4,] 0.00000000 0.00000000 0.00000000 0.044768039 0.00000000 0.000000
##
##
##
##
[12,] 0.00000000 0.00000000 0.01186178 0.000000000 0.01691718 0.000000
##
  [,575]
     [,576]
       [,577]
         [,578]
            [,579]
##
              [,580]
##
[2,] 0.00000000 0.00000000 0.00000000 0.01000838 0.005887281 0.00000000
##
##
##
##
[12,] 0.01186178 0.00000000 0.01186178 0.00000000 0.000000000 0.04671448
##
  [,581]
     [,582]
       [,583]
         [,584]
            [,585]
              [,586]
```

```
[1,] 0.07698643 0.00000000 0.00000000 0.00000000 0.01421308 0.00000000
 [2,] 0.01146508 0.01429768 0.00000000 0.00000000 0.00000000 0.01112042
##
 ##
##
 ##
 ##
 [11,] 0.00000000 0.00000000 0.04448168 0.00000000 0.00000000 0.00000000
[12,] 0.01930960 0.00000000 0.00000000 0.13468906 0.00000000 0.00000000
[13,] 0.00000000 0.00000000 0.00000000 0.03666604 0.00000000 0.00000000
##
##
          [,589]
    [,587]
       [,588]
             [,590]
                 [,591]
##
 [1,] 0.000000000 0.02896405 0.00000000 0.09211753 0.00000000 0.000000000
 [2,] 0.000000000 0.01934960 0.02327278 0.00000000 0.01251047 0.008564321
##
 ##
 ##
 ##
 [12,] 0.008148009 0.06879857 0.00000000 0.00000000 0.00000000 0.010150306
 ##
##
   [,593]
       [,594]
          [,595]
             [,596]
                 [,597]
                    [,598]
##
 ##
 [2,] 0.01112042 0.01251047 0.005887281 0.00000000 0.00000000 0.00000000
 ##
 ##
 ##
 [14,] 0.00000000 0.00000000 0.000000000 0.02729557 0.00000000 0.00000000
##
      [,600]
          [,601]
             [,602]
                 [,603]
 [1,] 0.0000000 0.00000000 0.013593463 0.000000000 0.00000000 0.00000000
##
##
 [2,] 0.0000000 0.01112042 0.008999376 0.008340314 0.00000000 0.01270531
 ##
 ##
 [5,] 0.0000000 0.00000000 0.020461738 0.000000000 0.00000000 0.00000000
 ##
##
 ##
```

```
[13,] 0.0000000 0.00000000 0.032397753 0.000000000 0.00000000 0.00000000
##
[,605]
      [,606]
          [,607]
             [,608]
                [,609]
##
 [2,] 0.00000000 0.0000000 0.000000000 0.01527276 0.00000000 0.023562
##
##
 ##
 [6,] 0.00000000 0.0000000 0.000000000 0.00000000 0.00000000 0.0000000
##
 ##
 [12,] 0.00000000 0.0000000 0.008896335 0.01086063 0.00000000 0.028798
##
   [,611]
       [,612]
          [,613]
             [,614]
                [,615]
 ##
 [2,] 0.00000000 0.01018184 0.01668063 0.01018184 0.00000000 0.00000000
##
 [4,] 0.00000000 0.00000000 0.00000000 0.02290915 0.00000000 0.00000000
##
 ##
 [12,] 0.01186178 0.00000000 0.00000000 0.00000000 0.01617516 0.00000000
##
             [,620]
       [.618]
          [,619]
##
   [,617]
                 [,621]
                    [,622]
##
 [1,] 0.00000000 0.000000000 0.03738202 0.00000000 0.00000000 0.000000000
 [2,] 0.01527276 0.007636382 0.01822374 0.01429768 0.008999376 0.005090658
##
 ##
 ##
 ##
 ##
##
 [10,] 0.05430316 0.000000000 0.06171000 0.00000000 0.000000000 0.000000000
##
[,625]
       [,624]
##
   [,623]
             [,626]
                [,627]
                   [,628]
##
 [1,] 0.00000000 0.00000000 0.00000000 0.03413016 0.02136116 0.0000000
 [2,] 0.01112042 0.01000838 0.00000000 0.01331076 0.00000000 0.0238940
```

```
##
 [6,] 0.00000000 0.00000000 0.05864741 0.00000000 0.00000000 0.00000000
 ##
##
 ##
 [9,] 0.00000000 0.00000000 0.00000000 0.05280303 0.00000000 0.00000000
 [14,] 0.00000000 0.00000000 0.00000000 0.06252025 0.07608563 0.0000000
       [,630]
          [,631]
             [,632]
                 [,633]
##
   [,629]
##
 [1,] 0.0000000 0.00000000 0.0000000 0.02375262 0.00000000 0.00000000
 ##
 ##
 [4,] 0.0000000 0.00000000 0.00000000 0.02105346 0.00000000 0.00000000
 ##
 ##
 ##
 [11,] 0.0000000 0.02819587 0.0000000 0.05146401 0.00000000 0.18318145
 [12.] 0.0000000 0.00000000 0.0000000 0.02058560 0.00000000 0.05587234
 [13,] 0.0000000 0.00000000 0.3748592 0.00000000 0.00000000 0.00000000
 [,635]
       [,636]
           [,637]
              [,638]
                  [,639]
                     [,640]
##
 ##
 [2,] 0.01112042 0.01668063 0.00000000 0.00000000 0.01309094 0.01112042
##
 ##
 [4,] 0.00000000 0.00000000 0.00000000 0.04790094 0.00000000 0.00000000
##
 ##
 [8,] 0.00000000 0.00000000 0.04448168 0.00000000 0.00000000 0.00000000
 ##
 [,641]
       [,642]
           [,643]
              [,644]
                  [,645]
                     [,646]
##
 [1,] 0.00000000 0.00000000 0.00000000 0.05119524 0.05558105 0.00000000
##
 ##
 ##
 [5,] 0.00000000 0.08006702 0.00000000 0.00000000 0.01856828 0.00000000
 ##
## [12,] 0.01186178 0.00000000 0.01186178 0.00000000 0.02842304 0.01086063
```

```
[14,] 0.00000000 0.00000000 0.00000000 0.16235097 0.07547148 0.00000000
##
           [,648]
                [,649] [,650]
 [1,] 0.00000000 0.00000000 0.00000000 0.02053 0.00000000 0.01246067
##
##
 [2,] 0.00000000 0.00000000 0.03639836 0.00000 0.01145457 0.00000000
 ##
 [4.] 0.00000000 0.04476804 0.00000000 0.00000 0.00000000 0.09326626
 [5,] 0.05498195 0.00000000 0.00000000 0.07330927 0.00000000
##
 [,657]
##
      [,653]
           [,654]
                 [,655]
                      [,656]
 ##
 [2,] 0.01427387 0.00000000 0.005887281 0.00000000 0.02001675 0.00000000
##
 ##
 ##
 [10,] 0.05075153 0.00000000 0.000000000 0.06981835 0.00000000 0.00000000
 [12,] 0.00000000 0.00000000 0.000000000 0.01810105 0.00000000 0.03599750
 ##
           [,660]
                [,661]
                     [,662]
                           [,663]
 [1,] 0.00000000 0.00000000 0.00000000 0.01567185 0.00000000 0.000000000
##
 [2,] 0.00000000 0.00000000 0.01251047 0.01426579 0.01251047 0.009163658
 ##
 [5,] 0.00000000 0.00000000 0.00000000 0.12626151 0.00000000 0.000000000
##
 [6,] 0.00000000 0.00000000 0.00000000 0.03056011 0.00000000 0.000000000
 ##
 [8,] 0.09309113 0.05719073 0.00000000 0.06063514 0.00000000 0.000000000
 [9,] 0.00000000 0.00000000 0.00000000 0.03750559 0.00000000 0.000000000
 [10.] 0.00000000 0.00000000 0.00000000 0.03395568 0.00000000 0.000000000
 [11,] 0.00000000 0.00000000 0.00000000 0.08731460 0.00000000 0.000000000
 [12,] 0.00000000 0.00000000 0.00000000 0.03131468 0.00000000 0.000000000
 [13,] 0.00000000 0.00000000 0.00000000 0.21799545 0.00000000 0.000000000
 ##
      [,665]
##
           [,666]
 [1,] 0.00000000 0.01756784
##
 [2,] 0.00000000 0.01070540
 [3,] 0.0000000 0.00000000
##
##
 [4,] 0.00000000 0.00000000
 [5.] 0.00000000 0.00000000
 [6,] 0.00000000 0.00000000
```

```
[7,] 0.00000000 0.00000000
   [8,] 0.00000000 0.00000000
##
  [9,] 0.00000000 0.00000000
## [10,] 0.00000000 0.00000000
## [11,] 0.00000000 0.00000000
## [12,] 0.00000000 0.02537576
## [13,] 0.00000000 0.00000000
## [14,] 0.02729557 0.00000000
##
## Cluster size:
   [1] 26 48 14 16 10 15 13 12 11 9 12 27 15 16
## Within-cluster sum of squares:
                            28.99900 82.42294 28.31112 30.86420 24.20945
        59.63401 105.32237
       21.79380 34.80995 28.80422 29.61511 47.63987 44.32458 29.51425
   [8]
table(result_sclust)
## result_sclust
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14
## 26 48 14 16 10 15 13 12 11 9 12 27 15 16
```

#### Our K-way clustering with QR decomposition

```
source("lib/SpectralClustering.R")
## compute cosine similarities matrix
docsdissim <- cosSparse(t(dtm_train_tfidf))
## apply K-WAY SPECTRAL CLUSTERING
k <- length(unique(AKumar$AuthorID))
result_sQRclust <- specClusteringQR(as.matrix(docsdissim), k)</pre>
```

We can also using hierarchical clustering method under the cosine distance. The intention of using a different clustering method is just to let you know how to compare performance between various methods.

```
start.time <- Sys.time()
docsdissim <- cosSparse(t(dtm_train_tfidf))
rownames(docsdissim) <- c(1:nrow(dtm_train_tfidf))
colnames(docsdissim) <- c(1:nrow(dtm_train_tfidf))
#compute pairwise cosine similarities using cosSparse function in package qlcMatrix
h <- hclust(as.dist(docsdissim), method = "ward.D")
result_hclust <- cutree(h,length(unique(AKumar$AuthorID)))
end.time <- Sys.time()
time_hclust <- end.time - start.time
table(result_hclust)</pre>
```

```
## result_hclust
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14
## 44 29 39 22 18 16 16 12 13 7 6 8 5 9
```

# Step 4: Evaluation

To evaluate the performance of the method, it is required to calculate the degree of agreement between a set of system-output partitions and a set of true partitions. In general, the agreement between two partitions is measured for a pair of entities within partitions. The basic unit for which pair-wise agreement is assessed is a

pair of entities (authors in our case) which belongs to one of the four cells in the following table (Kang et at.(2009)):

Matching matrix for the agreement between two sets of clusters

		Gold standard clusters (G)	
		Match	Mismatch
Machine-generated clusters (M)	Match Mismatch	a c	b d

Let M be the set of machine-generated clusters, and G the set of gold standard clusters. Then, in the table, for example, a is the number of pairs of entities that are assigned to the same cluster in each of M and G. Hence, a and d are interpreted as agreements, and b and c disagreements. When the table is considered as a confusion matrix for a two-class prediction problem, the standard "Precision", "Recall", "F1", and "Accuracy" are defined as follows.

$$\begin{aligned} & \operatorname{Precision} = \frac{a}{a+b} \\ & \operatorname{Recall} = \frac{a}{a+c} \\ & \operatorname{F1} = \frac{2 \times \operatorname{Precision} \times \operatorname{Recall}}{\operatorname{Precision} + \operatorname{Recall}} \\ & \operatorname{Accuracy} = \frac{a+d}{a+b+c+d} \end{aligned}$$

```
source('lib/evaluation_measures.R')
matching_matrix_hclust <- matching_matrix(AKumar$AuthorID,result_hclust)</pre>
performance hclust <- performance statistics(matching matrix hclust)</pre>
matching_matrix_sclust <- matching_matrix(AKumar$AuthorID,result_sclust)</pre>
performance_sclust <- performance_statistics(matching_matrix_sclust)</pre>
# our result (awful for now -> accuracy of 38%)
matching_matrix_sQRclust <- matching_matrix(AKumar$AuthorID,result_sQRclust)
performance_sQRclust <- performance_statistics(matching_matrix_sQRclust)</pre>
performance_sQRclust
## $precision
## [1] 0.2258346
##
## $recall
## [1] 0.7765689
##
## $f1
## [1] 0.3499112
##
## $mcc
## [1] 0.04844667
##
## $accuracy
## [1] 0.382716
compare_df <- data.frame(method=c("sClust","hClust"),</pre>
                          precision=c(performance_sclust$precision, performance_hclust$precision),
                          recall=c(performance sclust$recall, performance hclust$recall),
                          f1=c(performance_sclust$f1, performance_hclust$f1),
                          accuracy=c(performance_sclust$accuracy, performance_hclust$accuracy),
```

```
time=c(time_sclust,time_hclust))
kable(compare_df,caption="Comparision of performance for two clustering methods",digits = 2)
```

Table 2: Comparision of performance for two clustering methods

method	precision	recall	f1	accuracy	time
sClust hClust	0.34 0.20		0.20 0.12		0.73319197 secs 0.01806307 secs