Web and Network Assignment 2

Aditi Sonkusare

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(igraph)
library(kableExtra)
library(tidyverse)
library(tidyr)
library(ggraph)
library(ggrepel)
set.seed(100)


# reading the graph
wordpairs<- read_graph(file="./WordPairs.txt",format="pajek")
wordpairs<- as.undirected(wordpairs)
wordpairs<- simplify(wordpairs)

cues <- read.table("./cue.txt", header = F, sep="\t", skip=4)

V(wordpairs)$cue<-cues[[1]]

#checking the diameter value
print(diameter(wordpairs, weights = NA))
```

```
## [1] 7
```

PART 1

Q1.1

```
# function to check whether the passed target 1 and target 2 are the cue words:
check_cue <- function(target1, target2){
  if(V(wordpairs)[target1]$cue & V(wordpairs)[target2]$cue){
    cat("Both the targets are cue words \n")
  }else{
    cat("Both words are not cue words \n")
    cat(target1, "cue = ", as.logical(V(wordpairs)[target1]$cue), "\n")
    cat(target2, "cue = ", as.logical(V(wordpairs)[target2]$cue), "\n")
```

```
  }
}
```

Q1.1 :- rand_walk_topic_network: Random walk can travel on a graph randomly by moving from one node to another which generates a series of nodes in a graph. It is frequently used to analyze a network's structure.

```
random_walk_topic_network <- function(wordpairs,target_node_names,
                                      steps, walks, mode, topn){
  list1 <- c()
  for (i in 1:walks){
    list1 <- c(list1, list(random_walk(wordpairs, target_node_names[1],
                                       steps, mode = mode)))
  }
  frequency_words1 <- head(sort(table(names(unlist(list1))), decreasing = TRUE), 100)
  unique_words1 <- names(frequency_words1)

  list2 <- c()
  for (i in 1:walks){
    list2 <- c(list2, list(random_walk(wordpairs, target_node_names[2],
                                       steps, mode = mode)))
  }
  frequency_words2 <- head(sort(table(names(unlist(list2))),
                                decreasing = TRUE), topn)
  unique_words2 <- names(frequency_words2)


  output = append(unique_words1, unique_words2)
  return(unique(output))

}
```

Q1.1 :- ppr_topic_network:

```
page_rank_ppr <- function(wordpairs, names, damping , topn){

  teleport_probs <- rep(0,vcount(wordpairs))
  teleport_probs[as.numeric(V(wordpairs)[names])]<-1/length(names)

  pr<- page_rank(wordpairs,  directed = F,
                 personalized=teleport_probs,
                 damping = damping)$vector

  top_n_pr <- order(pr, decreasing=TRUE)[1:topn]

  top_n_pr<-V(wordpairs)[top_n_pr]
  word_association_network <- induced.subgraph(wordpairs, top_n_pr)
  return(word_association_network)
}
```

Q1.2 :- CENTRALITY MEASURES:

```r
centralities = function(word_association_network){
  eccentricity <- eccentricity(word_association_network)
  #eccentricity <- na.omit(eccentricity[!names(eccentricity) %in% c(target_word1, target_word2)])
  eccentricity <- head(sort(eccentricity, decreasing = TRUE), 5)

  betweenness <- betweenness(word_association_network)
  #betweenness <- betweenness[!names(betweenness) %in% c(target_word1, target_word2)]
  betweenness <- head(sort(betweenness, decreasing = TRUE), 5)

  eigen_centrality <- eigen_centrality(word_association_network)$vector
  #eigen_centrality <- eigen_centrality[!names(eigen_centrality) %in% c(target_word1, target_word2)]
  eigen_centrality <- head(sort(eigen_centrality, decreasing = TRUE), 5)
  return(tibble(eccentricity = names(eccentricity),betweenness = names(betweenness),
                eigen_centrality = names(eigen_centrality)))
}
```

```r
# FIRST PAIR:

#the first pair of targets: target1 as cocaine and target 2 as addiction:
target1 <- "COCAINE"
target2 <- "ADDICTION"

# first checking whether the passed targets are cue words or not:
check_cue(target1, target2)
```

```
## Both the targets are cue words
```

```r
# FOR RANDOM WALK:
# calling random_walk function and passing steps, walks, modes and topn here:
output <- random_walk_topic_network(wordpairs, c(target1, target2), 3, 100, "all", 160)

#getting association of names in from the upper code:
association <- V(wordpairs)[name %in% output]

#generating induced subgraph of wordpairs and association
association_network1 <- induced.subgraph(wordpairs, association)

# calling the centrality function here to measure the centrality of this network
# which returns the top 5 centralities of each centrality measure:
centrality1 <- centralities(association_network1)

# creating a kable which returns all three measures stored:
tibble(centrality1, .name_repair = "minimal") %>%
  kbl( caption = "TABLE 1: Centralities for Random Walk for First Pair of Targets") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left")
```

Comments: We can see three columns showing each centrality measure for random walk. In which betweenness and eigen centrality is performing more better as compared to eccentricity. As most related words are seen in eigen centrality - all the five words are related to the first pair of words. Among five, three words are related in betweenness centrality and eccentricity is giving all five words which are not related to first pair of targets.

Q1.3

Table 1: TABLE 1: Centralities for Random Walk for First Pair of Targets

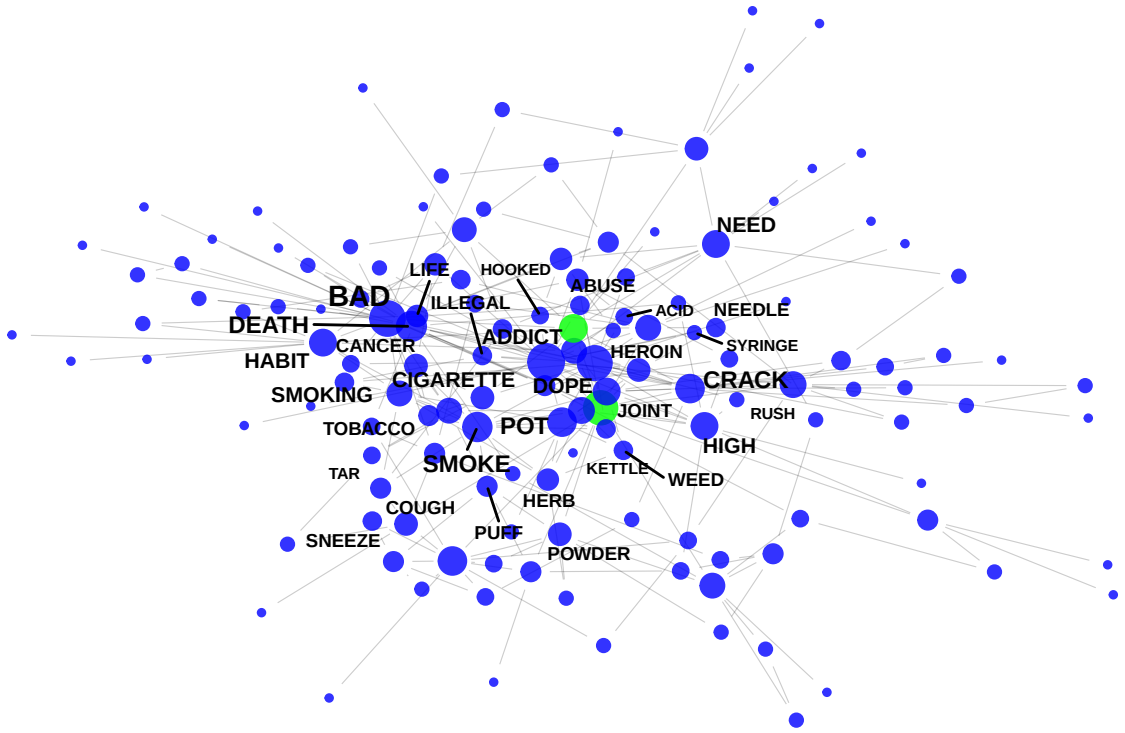
eccentricity	betweenness	eigen_centrality
LACK	COCAINE	CIGARETTE
RACE	BAD	SMOKE
SMART	DEATH	NICOTINE
HEAD	CRACK	DRUG
GET	DRUGS	TOBACCO

```
# visualisation for first pair of target for "random walk function"
vertex_size <- 2.5 + degree(association_network1)/1.5
cex_size <-2 + degree(association_network1)/12

labels <- names(sort(eigen_centrality(association_network1)$vector, decreasing = T)[1:40])

ggraph(association_network1, layout = "fr") +
  geom_edge_link(start_cap = circle(2.5, "mm"),
                   end_cap = circle(2.5, "mm"),
                   edge_width = 0.2,
                   alpha = 0.2) +
  geom_node_point(aes(size = vertex_size),
                    alpha = 0.8,
                    colour = ifelse(V(association_network1)$name %in% c(target1, target2),
                                    "#00FF00","#0000FF")) +
  labs(title = paste("Graph 1: Graph for Random Walk for Cocaine and Addiction")) +
  geom_node_text(
    aes(label = ifelse(V(association_network1)$name %in% labels, name, element_blank())),
    size = cex_size,
    fontface = "bold",
    repel = TRUE
    ) +
  theme(legend.position = "none",
        panel.background = element_rect(fill = "white"))
```

Table 2: TABLE 2: Centralities for Page Rank for First Pair of Targets

eccentricity	betweenness	eigen_centrality
SEX	COCAINE	CIGARETTE
BLACK	BAD	SMOKE
DUMB	DRUG	ASHTRAY
CAR	MEDICINE	CIGAR
WANT	GOOD	NICOTINE

Graph 1: Graph for Random Walk for Cocaine and Addiction




```
# FOR PAGE RANK FOR FIRST PAIR OF TARGET:
# calling personalised pagerank function and passing wordpairs, names, damping and topn here:
pg_output1 <- page_rank_ppr(wordpairs, c(target1, target2), 0.8, 160)

# calculating the centrality measures:
pg_centrality1 <- centralities(pg_output1)

#generating the kable which returns all the three measures stored:
tibble(pg_centrality1, .name_repair = "minimal") %>%
  kbl( caption = "TABLE 2: Centralities for Page Rank for First Pair of Targets") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left")
```

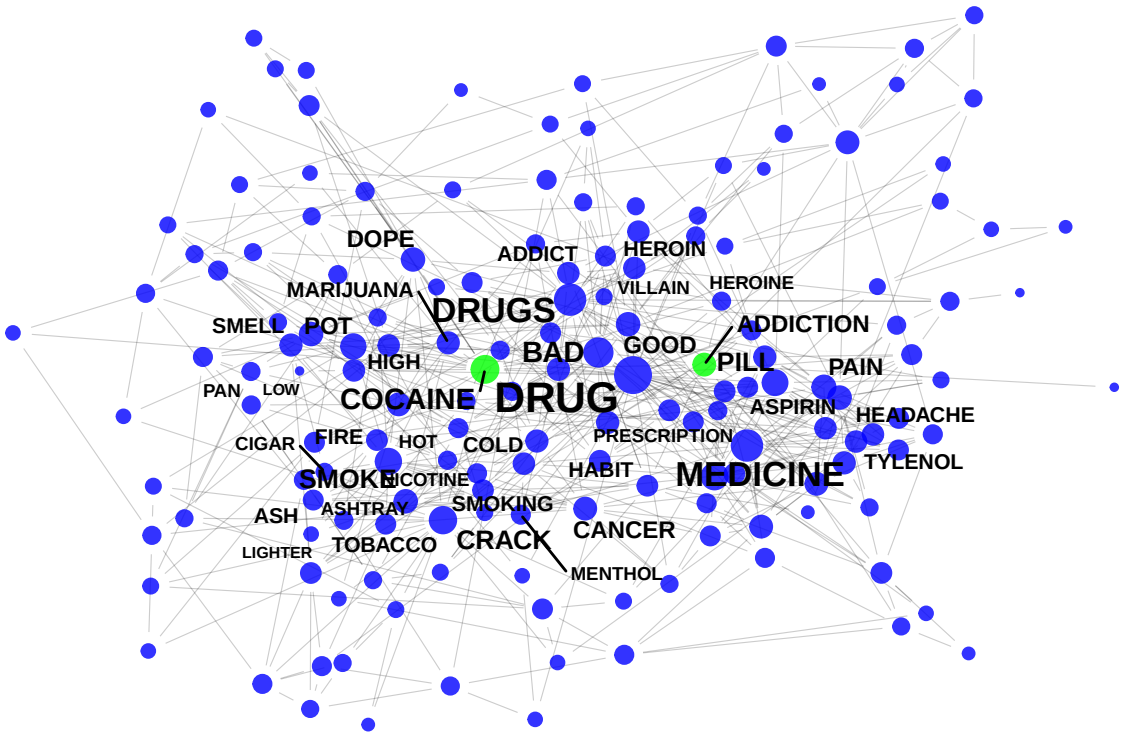
Comments: We can see three columns showing each centrality measure for page rank. In which betweenness and eigen centrality are providing more better as compared to eccentricity. As most related words are seen in eigen centrality - all the five words matches the first pair of words. Among five, three words are related

and two words are clearly not related in betweenness centrality and eccentricity is giving the words which are not related to first pair of targets.

Q1.3

```
# Visualisation for first set of words using "page rank algorithm"
pg_labels1 <- names(sort(eigen_centrality(pg_output1)$vector, decreasing = T)[1:40])

vertex_size <- 2.5 + degree(pg_output1)/1.5
cex_size <-2 + degree(pg_output1)/12

ggraph(pg_output1, layout = "fr") +
  geom_edge_link(start_cap = circle(2.5, "mm"),
                   end_cap = circle(2.5, "mm"),
                   edge_width = 0.2,
                   alpha = 0.2) +
  geom_node_point(aes(size = vertex_size),
                    alpha = 0.8,
                    colour = ifelse(V(pg_output1)$name %in% c(target1, target2),
                                    "#00FF00","#0000FF")) +
  labs(title = paste("Graph 2: Graph for Page Rank for Cocaine and Addiction")) +
  geom_node_text(
    aes(label = ifelse(V(pg_output1)$name %in% pg_labels1, name, element_blank())),
    size = cex_size,
    fontface = "bold",
    repel = TRUE
    ) +
  theme(legend.position = "none",
        panel.background = element_rect(fill = "white"))
```

Graph 2: Graph for Page Rank for Cocaine and Addiction




```
# SECOND PAIR:

#the second pair of targets: target1 as letter and target 2 as alphabet:
target1 <- "LETTER"
target2 <- "ALPHABET"

# first checking whether the passed targets are cue words or not:
check_cue(target1, target2)
```

```
## Both the targets are cue words
```

```
# FOR RANDOM WALK:
# calling random_walk function and passing steps, walks, modes and topn here:
output <- random_walk_topic_network(wordpairs, c(target1, target2), 3, 100, "all", 160)

#getting association of names in from the upper code:
association <- V(wordpairs)[name %in% output]

#generating induced subgraph of wordpairs and association
association_network2 <- induced.subgraph(wordpairs, association)

# calling the centrality function here to measure the centrality of this network
# which returns the top 5 centralities of each centrality measure:
centrality2 <- centralities(association_network2)
```

Table 3: TABLE 3: Centralities for Random Walk for Second Pair of Targets

eccentricity	betweenness	eigen_centrality
CHARGE	LETTER	LETTER
HAVE	ALPHABET	MAIL
FLY	SCHOOL	WRITE
MOVIE	GREEK	STAMP
AGGRESSIVE	WORD	ENVELOPE

```
# creating a kable which returns all three measures stored:
tibble(centrality2, .name_repair = "minimal") %>%
  kbl( caption = "TABLE 3: Centralities for Random Walk for Second Pair of Targets") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left")
```

Comments: We can see three columns showing each centrality measure for random walk, in which betweenness and eigen centrality is performing more better as compared to eccentricity. As betweenness centrality performed the best here - all five words are related to the second pair of target. Few words are related in eigen centrality which is not related and eccentricity is giving the words which are not related to second pair of targets.

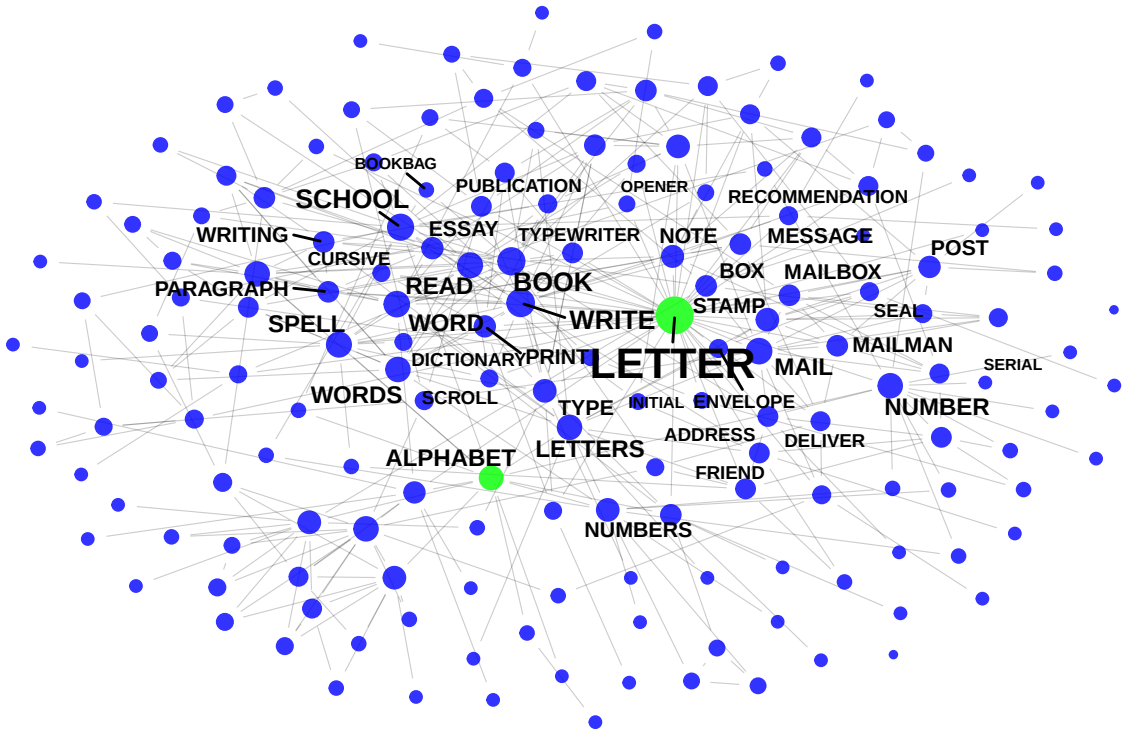
Q1.3

```
# visualisation for second pair of target for "random walk function"
vertex_size <- 2.5 + degree(association_network2)/1.5
cex_size <-2 + degree(association_network2)/12

labels <- names(sort(eigen_centrality(association_network2)$vector, decreasing = T)[1:40])

ggraph(association_network2, layout = "fr") +
  geom_edge_link(start_cap = circle(2.5, "mm"),
                   end_cap = circle(2.5, "mm"),
                   edge_width = 0.2,
                   alpha = 0.2) +
  geom_node_point(aes(size = vertex_size),
                    alpha = 0.8,
                    colour = ifelse(V(association_network2)$name %in% c(target1, target2),
                                    "#00FF00","#0000FF")) +
  labs(title = paste("Graph 3:  Graph for Random Walk for Letters and Alphabet")) +
  geom_node_text(
    aes(label = ifelse(V(association_network2)$name %in% labels, name, element_blank())),
    size = cex_size,
    fontface = "bold",
    repel = TRUE
    ) +
  theme(legend.position = "none",
        panel.background = element_rect(fill = "white"))
```

Table 4: TABLE 4: Centralities for Page Rank for Second Pair of Targets

eccentricity	betweenness	eigen_centrality
MANY	LETTER	WRITE
CHEESE	SCHOOL	LETTER
B	FOOD	PEN
ONE	MONEY	PAPER
ANIMAL	FRIEND	PENCIL

Graph 3: Graph for Random Walk for Letters and Alphabet



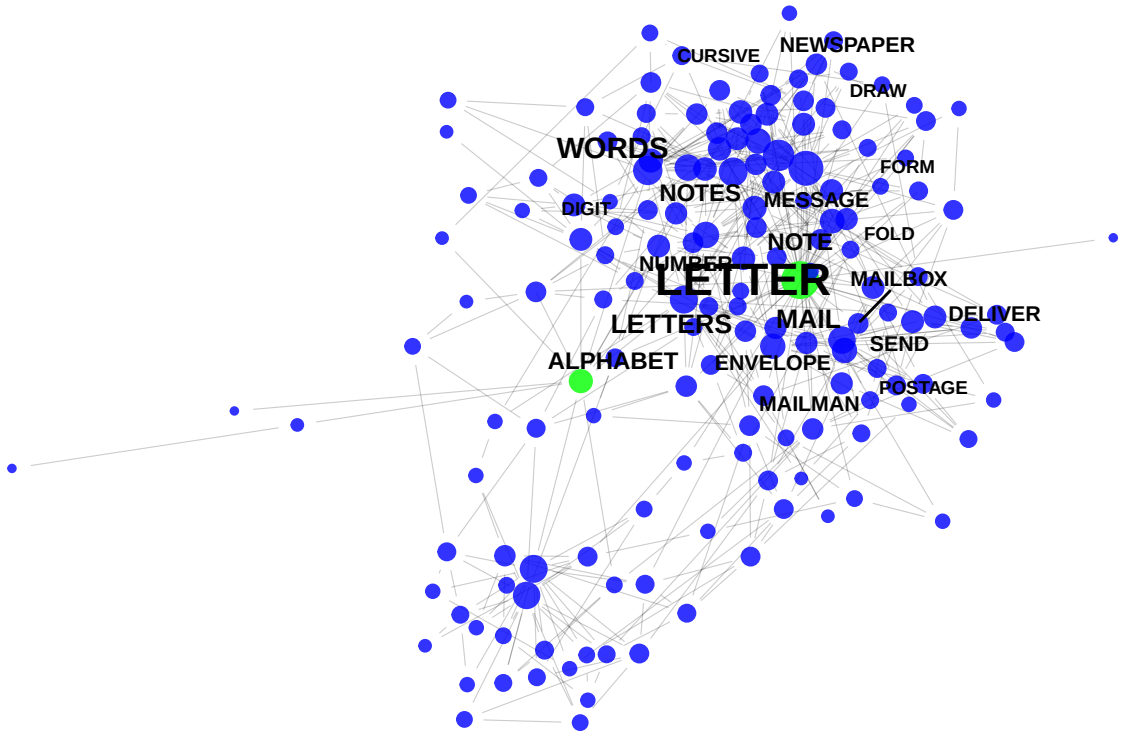

```
# FOR PAGE RANK FOR SECOND PAIR OF TARGET:
# calling personalised pagerank function and passing wordpairs, names, damping and topn here:
pg_output2 <- page_rank_ppr(wordpairs, c(target1, target2), 0.8, 160)

# calculating the centrality measures:
pg_centrality2 <- centralities(pg_output2)

#generating a kable which stored all the three measures:
tibble(pg_centrality2, .name_repair = "minimal") %>%
  kbl( caption = "TABLE 4: Centralities for Page Rank for Second Pair of Targets") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left")
```

Comments: We can see three columns showing each centrality measure for page rank, in which eigen centrality performed the best compared to eccentricity and betweenness for second pair of target. As all related words are seen in eigen centrality and very few words are related in betweenness centrality and eccentricity centrality.

Q1.3

```
# visualisation for second pair of target for "page rank function"

pg_labels2 <- names(sort(eigen_centrality(pg_output2)$vector, decreasing = T)[1:40])

vertex_size <- 2.5 + degree(pg_output2)/1.5
cex_size <-2 + degree(pg_output2)/12

ggraph(pg_output2, layout = "fr") +
  geom_edge_link(start_cap = circle(2.5, "mm"),
                 end_cap = circle(2.5, "mm"),
                 edge_width = 0.2,
                 alpha = 0.2) +
  geom_node_point(aes(size = vertex_size),
                   alpha = 0.8,
                   colour = ifelse(V(pg_output2)$name %in% c(target1, target2),
                                   "#00FF00","#0000FF")) +
  labs(title = paste("Graph 4: Graph for Page Rank for Letters and Alphabet")) +
  geom_node_text(
    aes(label = ifelse(V(pg_output2)$name %in% pg_labels2, name, element_blank())),
    size = cex_size,
    fontface = "bold",
    repel = TRUE
    ) +
  theme(legend.position = "none",
        panel.background = element_rect(fill = "white"))
```

Graph 4: Graph for Page Rank for Letters and Alphabet




```
# THIRD PAIR:

#the third pair of targets: target1 as missing and target 2 as kidnap:
target1 <- "MISSING"
target2 <- "KIDNAP"

# first checking whether the passed targets are cue words or not:
check_cue(target1, target2)
```

```
## Both the targets are cue words
```

```
# FOR RANDOM WALK:
# calling random_walk function and passing steps, walks, modes and topn here:
output <- random_walk_topic_network(wordpairs, c(target1, target2), 3, 100, "all", 160)

#getting association of names in from the upper code:
association <- V(wordpairs)[name %in% output]

#generating induced subgraph of wordpairs and association
association_network3 <- induced.subgraph(wordpairs, association)

# calling the centrality function here to measure the centrality of this network
# which returns the top 5 centralities of each centrality measure:
centrality3<- centralities(association_network3)
```

Table 5: TABLE 5: Centralities for Random Walk for Third Pair of Targets

eccentricity	betweenness	eigen_centrality
NOT HERE	CHILD	LOST
WITH	BAD	FOUND
ATTIC	LOVE	MISSING
TERRIBLE	PERSON	WANDER
SITUATION	GIRL	STOLEN

```
# creating a kable which returns all three measures stored:
tibble(centrality3, .name_repair = "minimal") %>%
  kbl( caption = "TABLE 5: Centralities for Random Walk for Third Pair of Targets") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left")
```

Comments: We can see three columns showing each centrality measure for random walk, in which eigen centrality performed the best - all words given by eigen are related. Betweenness also performed good - gave one word which is not related to the third pair of target. Eccentricity centrality giving word which are and which are not related to the third pair of target.

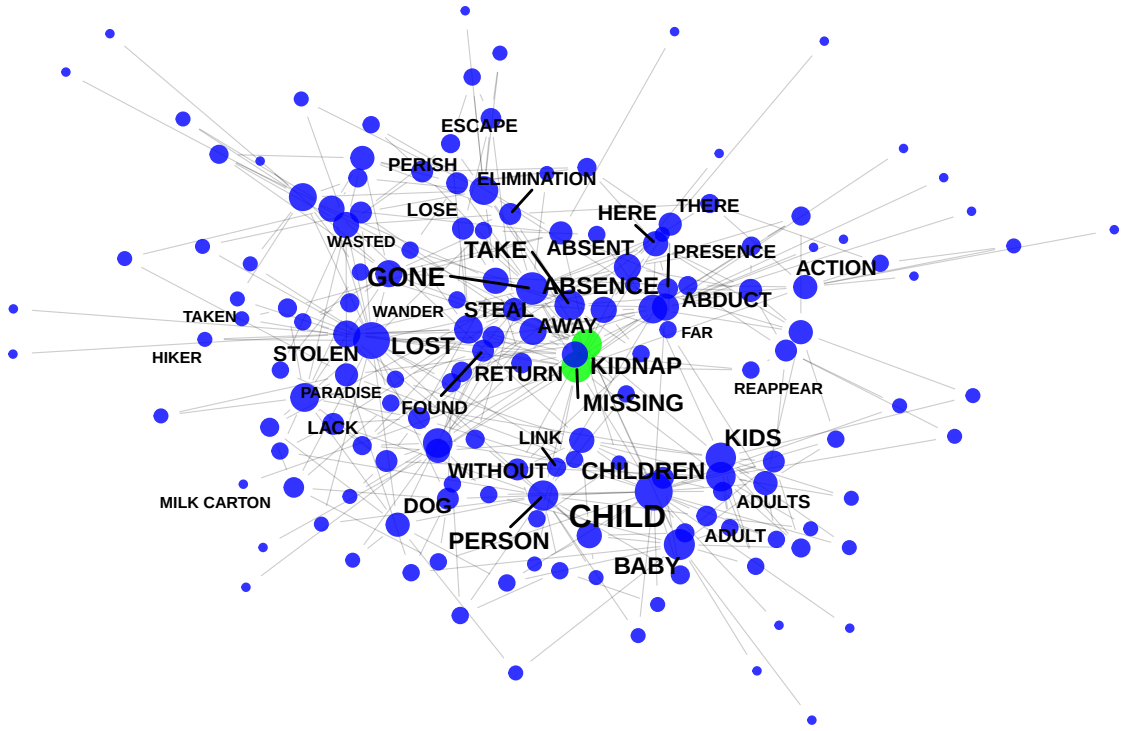
Q1.3

```
# visualisation for third pair of target for "random walk function"
vertex_size <- 2.5 + degree(association_network3)/1.5
cex_size <-2 + degree(association_network3)/12

labels <- names(sort(eigen_centrality(association_network3)$vector, decreasing = T)[1:40])

ggraph(association_network3, layout = "fr") +
  geom_edge_link(start_cap = circle(2.5, "mm"),
                 end_cap = circle(2.5, "mm"),
                 edge_width = 0.2,
                 alpha = 0.2) +
  geom_node_point(aes(size = vertex_size),
                  alpha = 0.8,
                  colour = ifelse(V(association_network3)$name %in% c(target1, target2),
                                  "#00FF00","#0000FF")) +
  labs(title = paste("Graph 5: Graph for Random Walk for Kidnap and Missing")) +
  geom_node_text(
    aes(label = ifelse(V(association_network3)$name %in% labels, name, element_blank())),
    size = cex_size,
    fontface = "bold",
    repel = TRUE
    ) +
  theme(legend.position = "none",
        panel.background = element_rect(fill = "white"))
```

Table 6: TABLE 6: Centralities for Page Rank for Third Pair of Targets

eccentricity	betweenness	eigen_centrality
LATE	MONEY	RIGHT
BURGLARY	STOLEN	LEFT
STEAL	GET	WRONG
ALIEN	GONE	TAKE
ADDUCT	GOOD	STEAL

Graph 5: Graph for Random Walk for Kidnap and Missing




```
# FOR PAGE RANK FOR THIRD PAIR OF TARGET:
# calling personalised pagerank function and passing wordpairs, names, damping and topn here:
pg_output3 <- page_rank_ppr(wordpairs, c(target1, target2), 0.8, 160)

# calculating the centrality measures:
pg_centrality3 <- centralities(pg_output3)

#generating a kable which stored all the three measures:
tibble(pg_centrality3, .name_repair = "minimal") %>%
  kbl( caption = "TABLE 6: Centralities for Page Rank for Third Pair of Targets") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left")
```

Comments: We can see three columns showing each centrality measure for page rank, in which all the three centralities are giving related and are also giving not related words. All three are performing nicely but the results are not the best as compared to previous results.

Q1.3

```
# visualisation for third pair of target for "random walk function"
pg_labels3 <- names(sort(eigen_centrality(pg_output3)$vector, decreasing = T)[1:40])

vertex_size <- 2.5 + degree(pg_output3)/1.5
cex_size <-2 + degree(pg_output3)/12

ggraph(pg_output3, layout = "fr") +
  geom_edge_link(start_cap = circle(2.5, "mm"),
                 end_cap = circle(2.5, "mm"),
                 edge_width = 0.2,
                 alpha = 0.2) +
  geom_node_point(aes(size = vertex_size),
                   alpha = 0.8,
                   colour = ifelse(V(pg_output3)$name %in% c(target1, target2),
                                   "#00FF00","#0000FF")) +
  labs(title = paste("Graph 6: Graph for Page Rank  for Kidnap and Missing")) +
  geom_node_text(
    aes(label = ifelse(V(pg_output3)$name %in% pg_labels3, name, element_blank())),
    size = cex_size,
    fontface = "bold",
    repel = TRUE
    ) +
  theme(legend.position = "none",
        panel.background = element_rect(fill = "white"))
```

Graph 6: Graph for Page Rank for Kidnap and Missing

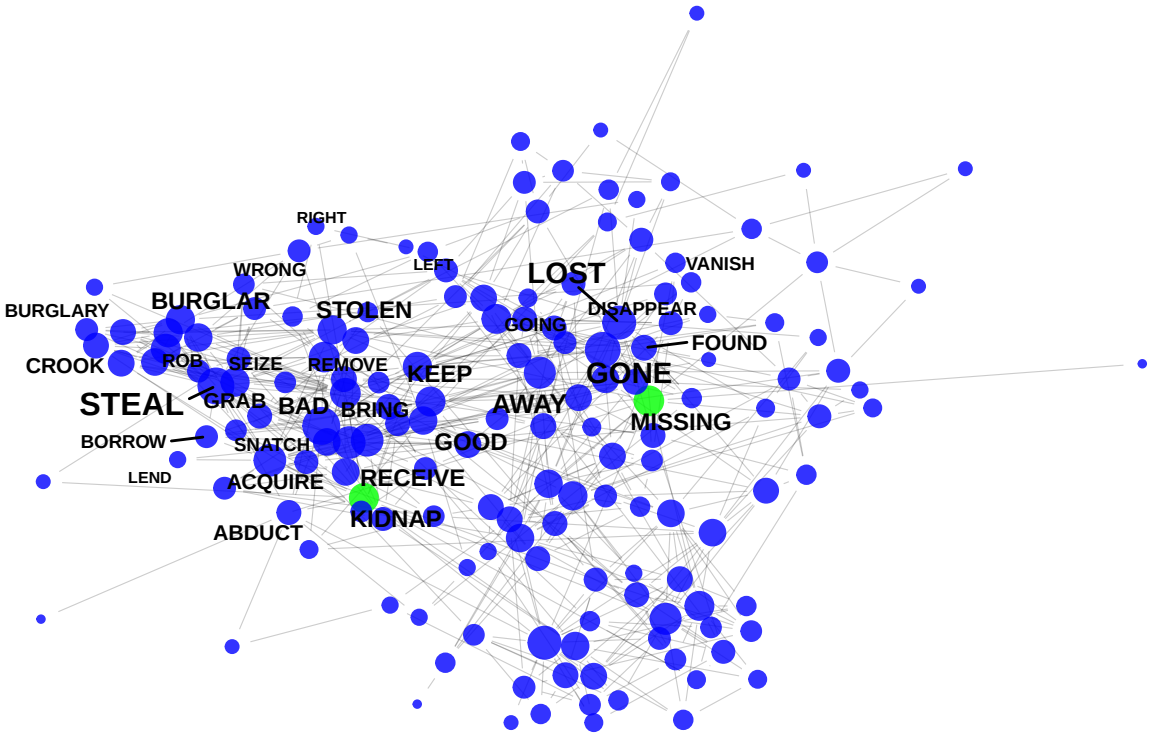


Table 7: TABLE 7: RANDOM WALK FOR ALL THREE PAIR OF TARGETS

COCAINE AND ADDICTION			LETTER AND ALPHABET			MISSING AND KIDNAP		
eccentricity	betweenness	eigen_centrality	eccentricity	betweenness	eigen_centrality	eccentricity	betweenness	eigen_centrality
LACK	COCAINE	CIGARETTE	CHARGE	LETTER	LETTER	NOT HERE	CHILD	LOST
RACE	BAD	SMOKE	HAVE	ALPHABET	MAIL	WITH	BAD	FOUND
SMART	DEATH	NICOTINE	FLY	SCHOOL	WRITE	ATTIC	LOVE	MISSING
HEAD	CRACK	DRUG	MOVIE	GREEK	STAMP	TERRIBLE	PERSON	WANDER
GET	DRUGS	TOBACCO	AGGRESSIVE	WORD	ENVELOPE	SITUATION	GIRL	STOLEN

Table 8: TABLE 8: PAGE RANK FOR ALL THREE PAIRS OF TARGETS

COCAINE AND ADDICTION			LETTER AND ALPHABET			MISSING AND KIDNAP		
eccentricity	betweenness	eigen_centrality	eccentricity	betweenness	eigen_centrality	eccentricity	betweenness	eigen_centrality
SEX	COCAINE	CIGARETTE	MANY	LETTER	WRITE	LATE	MONEY	RIGHT
BLACK	BAD	SMOKE	CHEESE	SCHOOL	LETTER	BURGLARY	STOLEN	LEFT
DUMB	DRUG	ASHTRAY	B	FOOD	PEN	STEAL	GET	WRONG
CAR	MEDICINE	CIGAR	ONE	MONEY	PAPER	ALIEN	GONE	TAKE
WANT	GOOD	NICOTINE	ANIMAL	FRIEND	PENCIL	ADDUCT	GOOD	STEAL

```
# Generating a single tibble to print all three centrality measures
# for the three pairs all-together for "random walk"

tibble(centrality1, centrality2, centrality3, .name_repair = "minimal") %>%
  kbl( caption = "TABLE 7: RANDOM WALK FOR ALL THREE PAIR OF TARGETS") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left",
                font_size = 6) %>%
  add_header_above(c("COCAINE AND ADDICTION" =3,"LETTER AND ALPHABET"= 3, "MISSING AND KIDNAP" = 3))
```

Comments: In conclusion to all three pairs of targets, there is atleast one centrality in each pair of target which is performing very poor. There is only one centrality measure among three which is best performing which is also giving the satisfactory results. The remaining one measure of centrality is giving related and not related words for which the results are poorly satisfactory on Random walk.

```
# Generating a single tibble to print all three centrality measures
# for the three pairs all-together for "page rank"

tibble(pg_centrality1, pg_centrality2, pg_centrality3, .name_repair = "minimal") %>%
  kbl(caption = "TABLE 8: PAGE RANK FOR ALL THREE PAIRS OF TARGETS") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left",font_size = 6 ) %>%
  add_header_above(c("COCAINE AND ADDICTION" =3,"LETTER AND ALPHABET"= 3, "MISSING AND KIDNAP" = 3))
```

Comments: In conclusion to all three pairs of targets, there are two centrality measures in each pair of target which is performing the best and there is only one centrality measure among three which is performing good but is failing to provide satisfactory results; which is giving related and not related words equally on page rank.

Finally concluding, as page rank gives two centrality measures which satisfies the pair of target passed and random walk only gives one centrality measure which satisfies the pair of target passed. Page rank is performing up to the mark and also giving satisfactory results as compared to random walk.

PART 2

Q2.1 :-

```
#FOR RANDOM WALK
#fast greedy community detection
RW_1 <- cluster_fast_greedy(association_network1)

#walktrap community detection
RW_2 <- cluster_fast_greedy(association_network2)

#fast greedy community detection
RW_3 <- cluster_infomap(association_network3)

#FOR PAGE RANK:
#walktrap community detection
PG_1 <- cluster_fast_greedy(pg_output1)

#walktrap community detection
PG_2 <- cluster_louvain(pg_output2)

#walktrap community detection
PG_3 <- cluster_edge_betweenness(pg_output3)


community_table <- function(c_data){
  s <- NULL
  l <- c()
  for(i in 1:length(c_data)){
    string = ""
    for(word in c_data[[i]]){
      string = paste0(string, "'", word, "',")
    }

    s <- c(s, string)
    l <- c(l, length(c_data[[i]]))
  }
  print(length(s))
  print(length(l))
  comm_table <- tibble(s, l)
  #arrange(desc(lengths))
  colnames(comm_table) <- c("C","S")
  comm_table <-  comm_table %>% arrange(desc(S))
  return(comm_table)
}

View(community_table(RW_1))
```

```
## [1] 13
## [1] 13
```

```
View(community_table(RW_2))
```

```
## [1] 14
## [1] 14
```

```
View(community_table(RW_3))
```

```
## [1] 25
## [1] 25
```

```
View(community_table(PG_1))
```

```
## [1] 14
## [1] 14
```

```
View(community_table(PG_2))
```

```
## [1] 14
## [1] 14
```

```
View(community_table(PG_3))
```

```
## [1] 10
## [1] 10
```

Comments: FOR ALGORITHMS ON RANDOM WALK:- 1. Out of other algorithms like walktrap, edge betweeness, infomap; fast greedy cluster performed the best on random walk for first pair of targets, Which gave the size of the cluster as 22. Here, the particular clusters are providing the output of very few not related words and most of the words in this are related words to the first pair of targets.

2. Out of other algorithms like label prop, louvain, infomap; fast greedy cluster performed the best on random walk for second pair of targets, Which gave the size of the cluster as 33 while other algorithms gave very high size of the particular cluster. Here, the particular clusters are providing the output of very few not related words and most of the words in this are related words to the second pair of targets.
3. Out of other algorithms; infomap cluster performed the best on random walk for third pair of targets, Which gave the size of the cluster as 17 while other algorithms gave very high size of the particular cluster. This gave the best result. Here, the particular clusters are providing the output of very few not related words and most of the words in this are related words to the third pair of targets.

FOR ALGORITHMS ON PAGE RANK:- 1. After comparing other algorithms like louvain, edge betweeness, infomap to fast greedy; fast greedy cluster performed the best on page rank for first pair of targets, Which gave the size of the cluster as 28 whereas, other algorithm gave very high cluster size of a particular cluster. Here, the particular clusters are providing the output of very few not related words and most of the words in this are related words to the first pair of targets.

2. After comparing other algorithms; louvain cluster performed the best on page rank for second pair of targets, Which gave the size of the cluster as 29 while other algorithms gave very high size of the particular cluster. Here, the particular clusters are providing the output of very few not related words and most of the words in this are related words to the second pair of targets.
3. After comparing other algorithms like label prop, fast greedy, leiden to edge betweeness; edge betweeness cluster performed the best on page rank for third pair of targets, Which gave the size of the cluster as 28 while other algorithms gave very high size of the particular cluster. Here, the particular clusters are providing the output of very few not related words and most of the words in this are related words to the third pair of targets.

Table 9: TABLE 9: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER FOR RANDOM WALK FOR FIRST PAIR OF TARGETS

C	S	central_label
'ABUSE','DRUG','USE','HOLE','ACID','DANGEROUS','COCAINE','HEROIN','ADDICTION','JANE'S','CRACK','TYLENOL','ECSTASY','ALLIGATOR','RUSH','DEATH','DEVASTATING','FRACTURE','PUT DOWN','LEAK','HEROINE','ADDICTIVE','SUPERWOMAN',	22	DEVASTATING
'NOSE','HEAD','BALLOON','SNEEZE','COUGH','DOG','FACE','SNORE','GLUE','FLUTE','BLOW','TISSUE','SNOT','WIPE','GUARD','POWDER','SNIFF','WINDSHIELD','SNUFF',	19	WINDSHIELD
'BAD','HABIT','ENVIRONMENT','SPELL','PLEASANT','NERVOUS','HANDCUFFS','BEHAVIOR','LIAR','NUN','ROUTINE','SPOIL','UNSTOPPABLE','TREND','UNPLEASANT',	15	UNPLEASANT
'ILLNESS','DANGER','AIDS','DEATH','KILL','LIFE','INSTINCT','SYMPATHY','CANCER','LOS ANGELES','COMA','SENTENCE','MISSILE','MAGGOT',	14	POT
'RACE','WHITE','CAGE','TIGER','WIRE','CLOTHES','COLOR','BLUSH','WHALE','BRACES','SWAN','GEESE','FRAY',	13	STUPID
'HIGH','IMPACT','TALL','JOINT','POT','HERB','CLIFF','MARIJUANA','GARLIC','KETTLE','WEED',	11	SNEEZE
'ADDICT','DRUGS','HOOKED','NO','PROHIBIT','LOSER','ILLEGAL','SEIZE','DOPE','PHARMACY',	10	BAD
'LACK','NEED','LOVE','DRIVE','EAT','DEMAND','COMPULSION','IMPULSE','THERAPY','PROVISION',	10	DRUGS
'STUPID','SMART','GET','BRAVE','OBNOXIOUS','FOOLISH','RIDICULE','CONCEIT','SHALLOW',	9	NEED
'CIGARETTE','SMOKE','TAR','TOBACCO','SMOKING','PUFF','NICOTINE',	7	DILEMMA
'PROBLEM','CONFUSION','DECIDE','PERSONAL','DILEMMA',	5	CIGARETTE
'WOMAN','LADY',	2	WOMAN
'NEEDLE','SYRINGE',	2	NEEDLE

Q2.2 and Q2.3 :-

```
library(tibble)
library(tidyverse)
library(kableExtra)
library(tibble)
lables_community <- function(cluster, table){
  central_label <- c()
  for(i in 1:length(cluster)){
    graph <- induced.subgraph(wordpairs, V(wordpairs)[cluster[[i]]])

    central_label <- c(central_label,
                       names(sort(eigen_centrality(graph)$vector, decreasing = T)[1]))
  }

  table <- tibble(table, central_label) %>% arrange(desc(S))

  return(table)
}
```

```
output_RW_1 <- lables_community(RW_1, community_table(RW_1))
```

```
## [1] 13
## [1] 13
```

```
tibble(output_RW_1, .name_repair = "minimal") %>%
  kbl(caption= "TABLE 9: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER
      FOR RANDOM WALK FOR FIRST PAIR OF TARGETS") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 4) %>%
  column_spec(column = 1, width = "5in")
```

Q2.4

```
output_RW_1 <- output_RW_1 %>%
  mutate(
    Interpretation = c(
       "Related to drug, cocain , addiction and dangerous which are a part of the target words",
       "Related to powder and sniff which are a part of the target words",
       "Related to habit and nervous which is a process to gain the target words",
```

Table 10: TABLE 10: INTERPRETATION FOR THE CLUSTER

C	S	central_label	Interpretation
'ABUSE','DRUG','USE','HOLE','ACID','DANGEROUS','COCAINE','HEROIN','ADDICTION','JANE'S','CRACK','TYLENOL','ECSTASY','ALLIGATOR','RUSH','DOPE','CRAVING','FRACTURE','PUFF', DOWN','LEAK','HEROINE','ADDICTIVE','SUPERWOMAN',	22	DOPE	Related to drug, cocain , addiction and d
'NOSE','HEAD','BALLOON','SNEEZE','COUGH','DOG','FACE','SNORE','GLUE','FLUTE','BLOW','TISSUE','SNOT','WIPE','GUARD','POWDER','SNIFF','WINDSHIELD','SNUFF',	19	WINDSHIELD	Related to powder and sniff which are a p
'BAD','HABIT','ENVIRONMENT','SPELL','PLEASANT','NERVOUS','HANDCUFFS','BEHAVIOR','LIAR','NUN','ROUTINE','SPOIL','UNSTOPPABLE','TEND','DEATH','PLEASANT',	17	DEATH	Related to habit and nervous which is a p
'ILLNESS','DANGER','AIDS','DEATH','KILL','LIFE','INSTINCT','SYMPATHY','CANCER','LOS ANGELES','COMA','SENTENCE','MISSILE','MAGGOT',	14	POT	Related to danger and death which are th
'RACE','WHITE','CAGE','TIGER','WIRE','CLOTHES','COLOR','BLUSH','WHALE','BRACES','SWAN','GEESE','FRAY',	13	STUPID	Non Relevant
'HIGH','IMPACT','TALL','JOINT','POT','HERB','CLIFF','MARIJUANA','GARLIC','KETTLE','WEED',	11	SNEEZE	Related to marijuna, high and herb are a
'ADDICT','DRUGS','HOOKED','NO','PROHIBIT','LOSER','ILLEGAL','SEIZE','DOPE','PHARMACY',	10	BAD	Related to addict and drugs clearly are a
'LACK','NEED','LOVE','DRIVE','EAT','DEMAND','COMPULSION','IMPULSE','THERAPY','PROVISION',	10	DRUGS	Non Relevant
'STUPID','SMART','GET','BRAVE','OBNOXIOUS','FOOLISH','RIDICULE','CONCEIT','SHALLOW',	9	NEED	Non Relevant
'CIGARETTE','SMOKE','TAR','TOBACCO','SMOKING','PUFF','NICOTINE',	7	DILEMMA	Related to smoke and tobacco are the ing
'PROBLEM','CONFUSION','DECIDE','PERSONAL','DILEMMA',	5	CIGARETTE	Non Relevant
'WOMAN','LADY',	2	WOMAN	Non Relevant
'NEEDLE','SYRINGE',	2	NEEDLE	Related to needles and syringes are a inst

```
      "Related to danger and death which are the post outcomes to the target pairs",
      "Non Relevant",
      "Related to marijuna, high and herb are a part of target pairs",
      "Related to addict and drugs clearly are a part of the target pairs",
      "Non Relevant",
      "Non Relevant",
      "Related to smoke and tobacco are the ingredients of the target pairs",
      "Non Relevant",
      "Non Relevant",
      "Related to needles and syringes are a instrument for the target pairs"
    )
  )

output_RW_1 %>%
  kbl(caption = "TABLE 10: INTERPRETATION FOR THE CLUSTER") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 4) %>%
  column_spec(column = 1, width = "5in")
```

```
#output_RW_1
View(output_RW_1)
```

Comments: Here, we can see the clusters, size of the cluster and one word label for the community. The community's label are stored in the central_label for each cluster. We can notice there are non relevant words in the cluster as the size of the respective cluster reduces. Higher the size of the cluster higher is the probability that the words are related to target pairs. Few clusters are non relevant, where the random walk model is performing good but not giving satisfactory results on random walk for first pair of target.

```
output_RW_2 <- lables_community(RW_2, community_table(RW_2))
```

```
## [1] 14
## [1] 14
```

```
tibble(output_RW_2, .name_repair = "minimal") %>%
  kbl(caption= "TABLE 11: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER
      FOR RANDOM WALK FOR SECOND PAIR OF TARGETS") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 4) %>%
  column_spec(column = 1, width = "5in")
```

Q2.4

Table 11: TABLE 11: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER FOR RANDOM WALK FOR SECOND PAIR OF TARGETS

C	S	central_label
'ALPHABET','FOOD','TRAGEDY','NUMBERS','FLY','SCARLET','ORDER','LETTERS','SOUP','ABC','KINDERGARTEN','CEREAL','TWENTY SIX','RUSH','GREEK','GREECE','BARLEY','WHEAT','BOWL','REQUEST','BOUILLON','SPOON','SALAD','SAUCER','FIBER','TOMATO','MUSHROOM','CRUNCHY','CORN FLAKES','PLAY DOUGH','PAPERS','CRUNCH','MYTH','NUMEROUS','INTAKE','SYMBOLS',	36	LETTER
'LETTER','RECEIVE','HAVE','MESSAGE','MAIL','PROMOTION','POST','FIRST','NAVY','ANNOUNCE','STAMP','BOX','NOTE','OPENER','BRING','DELIVERY','SEAL','MAILMAN','PUT','ELEVATOR','GARAGE','EN OFFICE','PARCEL','RECOMMENDATION','REFERRAL','WHOM',	36	LETTER
'WORD','DICTIONARY','VERB','ENGLISH','PRONOUN','GRAMMAR','AGGRESSIVE','WORDS','SPELL','BEE','WRITE','BOLD','TYPE','PRINT','BOUND','BOOK','WRITING','COMMA','SPEECH','ESSAY','CURSI	20	BOOK
'NUMBER','ADDRESS','HOME','PHONE','RESIDENCE','FRIEND','RING','BILL','NICE','COUNTER','DIAL','LOTTERY','TWO','GOSSIP','POPULAR','FIVE','WRITER','SERIAL','POPULATION',	19	WRITER
'SCHOOL','BOOK','MOVIE','READ','POLICY','EDITOR','PRINCIPAL','BOOKBAG','BEGINNING','CHARACTER','WITNESS','EQUATION','MONDAY','PUBLICATION'	14	ALPHABET
'NOTICE','NEWS','LOOK','INFORMATION','SURVEY','IMPORTANT','COMPUTER','BULLETIN','TYPEWRITER','EDITORIAL','NETWORK',	11	PLUS
'EMPTY','HEAD','MAGIC','FORM','DEPARTMENT','BLOCK','HAT','FILL',	8	SYMBOL
'A','THE','PLUS','AND','GREAT','ADD','ADDITION',	7	HEAD
'CHARGE','COVER','CARD','REPORT',	4	CARD
'NORTH','HEADING','DIRECTION','COURSE',	4	DIRECTION
'PATRIOTIC','SYMBOL','INFINITY',	3	NEWS
'EYES','GLASSES',	2	EYES
'HEART',	1	HEART
'INSINCERE',	1	INSINCERE

Table 12: TABLE 12: INTERPRETATION FOR THE CLUSTER

C	S	central_label	Explanation
'ALPHABET','FOOD','TRAGEDY','NUMBERS','FLY','SCARLET','ORDER','LETTERS','SOUP','ABC','KINDERGARTEN','CEREAL','TWENTY SIX','RUSH','GREEK','GREECE','BARLEY','WHEAT','BOWL','REQUEST','BOUILLON','SPOON','SALAD','SAUCER','FIBER','TOMATO','MUSHROOM','CRUNCHY','CORN FLAKES','PLAY DOUGH','PAPERS','CRUNCH','MYTH','NUMEROUS','INTAKE','SYMBOLS',	36	LETTER	Related to A, alphabet, number and lette
'LETTER','RECEIVE','HAVE','MESSAGE','MAIL','PROMOTION','POST','FIRST','NAVY','ANNOUNCE','STAMP','BOX','NOTE','OPENER','BRING','DELIVERY','SEAL','MAILMAN','PUT','ELEVATOR','GARAGE','EN OFFICE','PARCEL','RECOMMENDATION','REFERRAL','WHOM',	36	LETTER	[illegible]
'WORD','DICTIONARY','VERB','ENGLISH','PRONOUN','GRAMMAR','AGGRESSIVE','WORDS','SPELL','BEE','WRITE','BOLD','TYPE','PRINT','BOUND','BOOK','WRITING','COMMA','SPEECH','ESSAY','CURSI	20	BOOK	[illegible]
'NUMBER','ADDRESS','HOME','PHONE','RESIDENCE','FRIEND','RING','BILL','NICE','COUNTER','DIAL','LOTTERY','TWO','GOSSIP','POPULAR','FIVE','WRITER','SERIAL','POPULATION',	19	WRITER	[illegible] English and grammar w
'SCHOOL','BOOK','MOVIE','READ','POLICY','EDITOR','PRINCIPAL','BOOKBAG','BEGINNING','CHARACTER','WITNESS','EQUATION','MONDAY','PUBLICATION'	14	ALPHABET	Non Relevant
'NOTICE','NEWS','LOOK','INFORMATION','SURVEY','IMPORTANT','COMPUTER','BULLETIN','TYPEWRITER','EDITORIAL','NETWORK',	11	PLUS	Related to numbers, phone and percentag
'EMPTY','HEAD','MAGIC','FORM','DEPARTMENT','BLOCK','HAT','FILL',	8	SYMBOL	Related to articles and news somewhere r
'A','THE','PLUS','AND','GREAT','ADD','ADDITION',	7	HEAD	Related to names, headings and labels wh
'CHARGE','COVER','CARD','REPORT',	4	CARD	Related to computer and error somewhere
'NORTH','HEADING','DIRECTION','COURSE',	4	DIRECTION	Non Relevant
'PATRIOTIC','SYMBOL','INFINITY',	3	NEWS	Non Relevant
'EYES','GLASSES',	2	EYES	Related to give and take or return somew
'HEART',	1	HEART	Non Relevant
'INSINCERE',	1	INSINCERE	Non Relevant

```
output_RW_2 <- output_RW_2 %>%
  mutate(
    Explanation = c(
       "Related to A, alphabet, number and letters which are a part of target pairs",
       "Related to school, book, learn, hadwritings",
       "Related to message and letters which are related to target pairs",
       "Related to school, english and grammar which are related to target pairs",
       "Non Relevant",
       "Related to numbers, phone and percentages somewhere related to target pairs",
       "Related to articles and news somewhere related to target pairs",
       "Related to names, headings and labels which are a part of target pairs",
       "Related to computer and error somewhere related to target pairs",
       "Non Relevant",
       "Non Relevant",
       "Related to give and take or return somewhere related to target pairs",
       "Non Relevant",
       "Non Relevant"
     )
   )

output_RW_2 %>%
  kbl(caption = "TABLE 12: INTERPRETATION FOR THE CLUSTER") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 4) %>%
  column_spec(column = 1, width = "5in")
```

Table 13: TABLE 13: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER FOR RANDOM WALK FOR THIRD PAIR OF TARGETS

C	S	central_label
'CHILD','ABUSE','ADULT','BABY','GROWN','SMALL','SOFT','CARING','HELPER','RESPONSIBILITY','DOLL','REPRODUCE','CRAYON','COLT','JUNIOR','SCRIBBLE','NURSERY',	17	SCRIBBLE
'ABDUCT','KIDNAP','TAKE','STEAL','ALIEN','ADDUCT','HARASS','KIDNAPPER','OPEN','KEEP','BORROW','PUT','SNATCH','REMOVER',	14	CHILD
'ABSENCE','SICK','TARDY','ATTENDANCE','CLASS','EXCUSE','DISGUSTING','HIGH SCHOOL','MONDAY',	9	CATCH
'WITHOUT','LACK','TOGETHER','JOIN','WITH','BRACELET','LINK','WITHIN','RECRUIT',	9	LOST
'MISSING','LOST','HIKER','PARADISE','WANDER','FOUND','WASTED','MILK CARTON',	8	ATTENDANCE
'AWAY','ABSENT','GONE','NOT HERE','FAR','LOSE','DISSOLVE','ELIMINATION',	8	HERE
'KIDS','PLAY','LOVE','ADULTS','CHILDREN','PARENTS','BLOCKS','PLAYING',	8	AWAY
'BAD','TROUBLE','TERRIBLE','HORRIBLE','SITUATION','WORST','HAGAR','SUSPECT',	8	VOID
'DEATH','KILL','BLOOD','WAR','DIE','PREY','CEMETERY','PERISH',	8	CHILDREN
'SHOW','MOVIE','SCARY','ATTIC','MITTEN','SLIDE','REAPPEAR',	7	WITH
'JAIL','RANSOM','PRISONER','HOSTAGE','NEGOTIATION','IRAN','LEBANON',	7	MONEY
'CAPTURE','CATCH','POW','TRAP','ESCAPE','CHASE',	6	TERRIBLE
'MONEY','AMBITION','RAISE','PURSE','BOUNTY','GENEROUS',	6	ACTION
'FAST','STOP','ACTION','MOVEMENT','VERB','NEW YORK',	6	BACK
'PERSON','NAME','PLEASANT','MIDDLE','DEAF','WHO',	6	MOVIE
'MEAN','PUSHY','RUDE','SELFISH','RIDICULE','STINGY',	6	NAME
'PRESENCE','HERE','THERE','EXIST','THEIR',	5	GIRL
'MAD','DISCIPLINE','SCOLD','SPANK',	4	SPANK
'DOG','POUND','PRECIOUS','LAP',	4	MEAN
'STOLEN','BIKE','TAKEN','RIPPED',	4	DIE
'EMPTY','VOID','WRONG',	3	DOG
'VIETNAM','WHIP','TORTURE',	3	PRISONER
'BACK','RETURN',	2	STOLEN
'GIRL','GAL',	2	TORTURE
'CURSE','WITCH',	2	CURSE

```
View(output_RW_2)
```

Comments: Here, we can see the clusters, size of the cluster and one word label for the community. The community's label are stored in the central_label for each cluster. We can notice there are non relevant words in the cluster as the size of the respective cluster reduces. Higher the size of the cluster higher is the probability that the words are related to target pairs. Here, the random walk is performing good by giving only 3-4 as non relevant to the second target pairs.

```
output_RW_3 <- lables_community(RW_3, community_table(RW_3))
```

```
## [1] 25
## [1] 25
```

```
tibble(output_RW_3, .name_repair = "minimal") %>%
  kbl(caption= "TABLE 13: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER
      FOR RANDOM WALK FOR THIRD PAIR OF TARGETS") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 4) %>%
  column_spec(column = 1, width = "5in")
```

Q2.4

```
output_RW_3 <- output_RW_3 %>%
  mutate(
    Explanation = c(
       "Related to small and child related to target pairs",
       "Related to abduct, kidnap and steal which are clearly a part of target pairs",
       "Related to money, account and poor which are the outcomes of the target pairs",
       "Non Relevant",
       "Related to bad and horror are the process in the target pair",
       "Related to killer and scarry are the process in the target pair",
       "Related to missing, lost and find are a part and also the process of the target pair",
       "Related to lost and disappear are clearly a part of target pairs",
       "Related to absent and classes are clearly a part of target pairs",
```

Table 14: TABLE 14: INTERPRETATION FOR THE CLUSTER

C	S	central_label	Explanation
'CHILD','ABUSE','ADULT','BABY','GROWN','SMALL','SOFT','CARING','HELPER','RESPONSIBILITY','DOLL','REPRODUCE','CRAYON','COLT','JUNIOR','SCRIBBLE','NURSERY',	17	CHILD	Related to small and child related to ta
'ABDUCT','KIDNAP','TAKE','STEAL','ALIEN','ADDUCT','HARASS','KIDNAPPER','OPEN','KEEP','BORROW','PUT','SNATCH','REMOVER',	14	CHILD	Related to abduct, kidnap and steal wh
'ABSENCE','SICK','TARDY','ATTENDANCE','CLASS','EXCUSE','DISGUSTING','HIGH SCHOOL','MONDAY',	9	CATCH	Related to money, account and poor wh
'WITHOUT','LACK','TOGETHER','JOIN','WITH','BRACELET','LINK','WITHIN','RECRUIT',	9	LOST	Non Relevant
'MISSING','LOST','HIKER','PARADISE','WANDER','FOUND','WASTED','MILK CARTON',	8	ATTENDANCE	Related to bad and horror are the proc
'AWAY','ABSENT','GONE','NOT HERE','FAR','LOSE','DISSOLVE','ELIMINATION',	8	HERE	Related to killer and scarry are the pro
'KIDS','PLAY','LOVE','ADULTS','CHILDREN','PARENTS','BLOCKS','PLAYING',	8	AWAY	Related to missing, lost and find are a
'BAD','TROUBLE','TERRIBLE','HORRIBLE','SITUATION','WORST','HAGAR','SUSPECT',	8	VOID	Related to lost and disappear are clear
'DEATH','KILL','BLOOD','WAR','DIE','PREY','CEMETERY','PERISH',	8	CHILDREN	Related to absent and classes are clearl
'SHOW','MOVIE','SCARY','ATTIC','MITTEN','SLIDE','REAPPEAR',	7	WITH	Non Relevant
'JAIL','RANSOM','PRISONER','HOSTAGE','NEGOTIATION','IRAN','LEBANON',	7	MONEY	Related to emptyness and alone
'CAPTURE','CATCH','POW','TRAP','ESCAPE','CHASE',	6	TERRIBLE	Related to kill, torchure and attack are
'MONEY','AMBITION','RAISE','PURSE','BOUNTY','GENEROUS',	6	ACTION	Non Relevant
'FAST','STOP','ACTION','MOVEMENT','VERB','NEW YORK',	6	BACK	Related to capture or kidnap are clearl
'PERSON','NAME','PLEASANT','MIDDLE','DEAF','WHO',	6	MOVIE	Non Relevant
'MEAN','PUSHY','RUDE','SELFISH','RIDICULE','STINGY',	6	NAME	Non Relevant
'PRESENCE','HERE','THERE','EXIST','THEIR',	5	GIRL	Non Relevant
'MAD','DISCIPLINE','SCOLD','SPANK',	4	SPANK	Non Relevant
'DOG','POUND','PRECIOUS','LAP',	4	MEAN	Related to not present that is absent
'STOLEN','BIKE','TAKEN','RIPPED',	4	DIE	Related to steal and fear are a part of
'EMPTY','VOID','WRONG',	3	DOG	Related to stole are clearly a part of ta
'VIETNAM','WHIP','TORTURE',	3	PRISONER	Non Relevant
'BACK','RETURN',	2	STOLEN	Non Relevant
'GIRL','GAL',	2	TORTURE	Non Relevant
'CURSE','WITCH',	2	CURSE	Non Relevant

```
      "Non Relevant",
      "Related to emptyness and alone",
      "Related to kill, torchure and attack are a part of target pairs",
      "Non Relevant",
      "Related to capture or kidnap are clearly a part of target pairs",
      "Non Relevant",
      "Non Relevant",
      "Non Relevant",
      "Non Relevant",
      "Related to not present that is absent ",
      "Related to steal and fear are a part of target pairs",
      "Related to stole are clearly a part of target pairs",
      "Non Relevant",
      "Non Relevant",
      "Non Relevant",
      "Non Relevant"
    )
  )

output_RW_3 %>%
  kbl(caption = "TABLE 14: INTERPRETATION FOR THE CLUSTER") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 4) %>%
  column_spec(column = 1, width = "5in")
```

```
View(output_RW_3)
```

Comments: Here, we can see the clusters, size of the cluster and one word label for the community. The community's label are stored in the central_label for each cluster. We can notice there are non relevant words in the cluster as the size of the respective cluster reduces. Higher the size of the cluster higher is the probability that the words are related to target pairs. Here, we can see when the size is reducing the model is not performing up to the mark and giving 5-6 outputs for the clusters as non relevant.

In conclusion to Random walk performance on the respective target pairs, the community detection algorithms are working good but still there are clusters which gives non relevant words as the output. Even though, the model is performing fairly good and also returning most outputs of the clusters which are related and completely relevant to the target pairs

Table 15: TABLE 15: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER FOR PAGE RANK FOR FIRST PAIR OF TARGETS

C	S	central_label
'PAIN','SICK','HURT','PAPER','HELP','ADVIL','HEADACHE','TYLENOL','ASPIRIN','MEDICINE','PILL','DISEASE','DEATH','KILL','GUN','DOCTOR','ANTIBIOTIC','INJECTION','POISON','REMEDY','MEDICATION','TAI	28	DOCTOR
'DRUG','HAPPY','HOLE','COCAINE','HEROIN','ADDICTION','JANE'S','CRACK','FIX','ECSTASY','RUSH','HERO','BREAK','SPLIT','EGG','KNUCKLE','CREMATE','SUPERMAN','FRACTURE','PRESCRIPTION','HE	21	HOLE
'FOOD','CRIME','ADDICT','DRUGS','JOINT','POT','COOK','EGGS','FLOWER','ILLEGAL','MARIJUANA','MUSHROOM','DOPE','LEGAL','PAN','KETTLE','WEED','HALLUCINATION',	15	WEED
'HOT','FIRE','COLD','SICKNESS','ASH','ASHTRAY','CIGARETTE','SMOKE','CIGAR','TOBACCO','LIGHTER','SMOKING','MENTHOL','NICOTINE',	14	PAIN
'NOSE','CHILD','ABUSE','MISTREAT','USE','BABY','BIRD','DOG','SMELL','FACE','BLOW','SUBSTANCE','POWDER','SNIFF',	14	ABUSE
'SEX','MONEY','WANT','TREE','GREEN','NEED','LOVE','DESIRE','COMPULSION','OBSESSION','NECESSARY','WALNUT',	12	WANT
'WATER','ALCOHOL','HOOKED','BEER','DRINK','DRUNK','CAFFEINE','COFFEE','FISH','COKE','PLUMBER',	11	BAD
'GOOD','BAD','ACID','CHEMICAL','DANGEROUS','HABIT','TENDENCY','VILLAIN','HARMFUL',	9	DRUGS
'BLACK','SPEED','CAR','STREET','ROAD','WHITE','PAVEMENT','TAR','SIDEWALK',	9	BLACK
'HIGH','HARD','LOW','WIRE','ROCK','STONED',	6	DRINK
'PROBLEM','ANSWER','SOLUTION','TROUBLE','DILEMMA',	5	PROBLEM
'DREAM','SLEEP','RELAX','ILLUSION','SEDATIVE',	5	SLEEP
'THREAD','NEEDLE','INJECTION','SYRINGE',	4	NEEDLE
'STUPID','DUMB','SMART','IDIOT',	4	DUMB

```
output_PG_1 <- lables_community(PG_1, community_table(PG_1))
```

```
## [1] 14
## [1] 14
```

```
tibble(output_PG_1, .name_repair = "minimal") %>%
  kbl(caption= "TABLE 15: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER
      FOR PAGE RANK FOR FIRST PAIR OF TARGETS") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 4) %>%
  column_spec(column = 1, width = "5in")
```

Q2.4

```
output_PG_1 <- output_PG_1 %>%
  mutate(
    Explanation = c(
       "Related to poison, doctor and death",
       "Related to drug, cocaine and addiction",
       "Related to drugs, marijuana, drugs, illegal",
       "Related to nicotines, smoking, tobacco,ashtray, cigar and ciggrates",
       "Related to powder, sniff, blow, nose and substances",
       "Related to money, green, need, obsession and desire",
       "Related to alcohol, beer and drinks",
       "Related to bad, chemical, dangerous, habit and harmful",
       "Related to tar, speed and black",
       "Related to high, stoned, low",
       "Related to dilemma and problems",
       "Related to dream, illusion and relax",
       "Related to needles, injection and syringe ",
       "Non Relevant"
     )
   )

output_PG_1 %>%
  kbl(caption = "TABLE 16: INTERPRETATION FOR THE CLUSTER") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 4) %>%
  column_spec(column = 1, width = "5in")
```

Table 16: TABLE 16: INTERPRETATION FOR THE CLUSTER

C	S	central_label	Explanation
'PAIN','SICK','HURT','PAPER','HELP','ADVIL','HEADACHE','TYLENOL','ASPIRIN','MEDICINE','PILL','DISEASE','DEATH','KILL','GUN','DOCTOR',[illegible]	[illegible]	[illegible]	[illegible]
'DRUG','HAPPY','HOLE','COCAINE','HEROIN','ADDICTION','JANE'S','CRACK','FIX','ECSTASY','RUSH','HERO','BREAK','SPLIT','EGG','KNUCKLE',[illegible]	[illegible]	[illegible]	[illegible]
'FOOD','CRIME','ADDICT','DRUGS','JOINT','POT','COOK','EGGS','FLOWER','ILLEGAL','MARIJUANA','MUSHROOM','DOPE','LEGAL','PAN','KETTLE','WEED','HALLUCINATION'	18	WEED	Related to drugs, marijuana, drugs, illega
'HOT','FIRE','COLD','SICKNESS','ASH','ASHTRAY','CIGARETTE','SMOKE','CIGAR','TOBACCO','LIGHTER','SMOKING','MENTHOL','NICOTINE',	14	PAIN	Related to nicotines, smoking, tobacco,as
'NOSE','CHILD','ABUSE','MISTREAT','USE','BABY','BIRD','DOG','SMELL','FACE','BLOW','SUBSTANCE','POWDER','SNIFF',	14	ABUSE	Related to powder, sniff, blow, nose and
'SEX','MONEY','WANT','TREE','GREEN','NEED','LOVE','DESIRE','COMPULSION','OBSESSION','NECESSARY','WALNUT',	12	WANT	Related to money, green, need, obsession
'WATER','ALCOHOL','HOOKED','BEER','DRINK','DRUNK','CAFFEINE','COFFEE','FISH','COKE','PLUMBER',	11	BAD	Related to alcohol, beer and drinks
'GOOD','BAD','ACID','CHEMICAL','DANGEROUS','HABIT','TENDENCY','VILLAIN','HARMFUL',	9	DRUGS	Related to bad, chemical, dangerous, hab
'BLACK','SPEED','CAR','STREET','ROAD','WHITE','PAVEMENT','TAR','SIDEWALK',	9	BLACK	Related to tar, speed and black
'HIGH','HARD','LOW','WIRE','ROCK','STONED',	6	DRINK	Related to high, stoned, low
'PROBLEM','ANSWER','SOLUTION','TROUBLE','DILEMMA',	5	PROBLEM	Related to dilemma and problems
'DREAM','SLEEP','RELAX','ILLUSION','SEDATIVE',	5	SLEEP	Related to dream, illusion and relax
'THREAD','NEEDLE','INJECTION','SYRINGE',	4	NEEDLE	Related to needles, injection and syringe
'STUPID','DUMB','SMART','IDIOT',	4	DUMB	Non Relevant

Table 17: TABLE 17: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER FOR PAGE RANK FOR SECOND PAIR OF TARGETS

C	S	central_label
'GRADE','GOOD','CLASS','SCHOOL','DRAW','WRITER','BOOK','PAPER','PICTURE','SPELL','FORM','PENCIL','READ','WRITE','FOLD','NOTEBOOK','PEN','BOLD','TYPE','PRINT','HEADING','WRITING','NOTI	[illegible]	[illegible]
'FOOD','WATER','HOT','VEGETABLES','BIRD','SOUP','CEREAL','CRACKERS','EGGS','CHEESE','FISH','CHICKEN','BEEF','BOWL','CUP','STEW','CRACKERS','SANDWICH','BOUILLON','CUBE','BROTH','SPOON	[illegible]	[illegible]
'LETTER','ANIMAL','CAR','MAN','MAIL','FRIEND','DOG','STAMP','BOX','CARD','LICK','STICKER','SEND','DELIVER','SEAL','PACKAGE','MAILMAN','ENVELOPE','MAILBOX','POST OFFICE','POSTAGE','PARCEL','WHOM',	[illegible]	[illegible]
'MUSIC','WORK','MONEY','COLLECT','CARDS','JOB','MESSAGE','PHONE','TALK','STATE','BILL','NOTE','TELEPHONE','CAPITAL','RECOMMENDATION'	[illegible]	LETTER
'ALPHABET','MATH','ADD','NUMBERS','CALCULATOR','LOVE','LETTERS','ABC','KINDERGARTEN','TWENTY SIX','ARTS','STATISTICS','PLAY DOUGH','PAPERS',	14	NUMBER
'MANY','NOTICE','TELEVISION','NEWSPAPER','NEWS','BOARD','BULLETIN','EDITORIAL','NUMEROUS','NETWORK','SERIES',	11	WORDS
'ONE','NUMBER','NAME','SIGN','FIRST','COUNT','SYMBOL','DIGIT','INITIAL','SERIAL',	10	JOB
'WORD','DICTIONARY','ENGLISH','VOCABULARY','WORDS','GREEK','FRATERNITY','SENTENCE','LANGUAGE','PARAGRAPH',	10	GIVE
'POST','OFFICE','COMPUTER','SECRETARY','TYPEWRITER','TYPIST',	6	OPENER
'HEAD','ADDRESS','HOME','HOUSE','BLOCK',	5	BULLETIN
'TAKE','GIVE','RECEIVE','GET',	4	SOUP
'A','B','VITAMIN',	3	HOME
'CAN','CLOSE','OPENER',	3	SECRETARY
'CHOWDER','CLAM',	2	CHOWDER

```
#output_RW_1
View(output_PG_1)
```

Comments: Here, we can see the clusters, size of the cluster and one word label for the community. The community's label are stored in the central_label for each cluster. As we can see that the model on Page Rank is performing very good and also giving satisfactory results and outputs, as there is only one clusters which is non relevant to the target pairs and other all are related to target pairs. As the size of the cluster is reducing still the page rank is giving related words to the target pair.

```
output_PG_2 <- lables_community(PG_2, community_table(PG_2))
```

```
## [1] 14
## [1] 14
```

```
tibble(output_PG_2, .name_repair = "minimal") %>%
  kbl(caption= "TABLE 17: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER
      FOR PAGE RANK FOR SECOND PAIR OF TARGETS") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 4) %>%
  column_spec(column = 1, width = "5in")
```

Q2.4

```
output_PG_2 <- output_PG_2 %>%
  mutate(
    Explanation = c(
       "Related to book, paper, write, notes, heading and type",
       "Not Related",
```

Table 18: TABLE 18: INTERPRETATION FOR THE CLUSTER

C	S	central_label	Explanation
'GRADE','GOOD','CLASS','SCHOOL','DRAW','WRITER','BOOK','PAPER','PICTURE','SPELL','FORM','PENCIL','READ','WRITE','FOLD','NOTEBOOK',	24	PEN	Related to grade, work, class, school and
'FOOD','WATER','HOT','VEGETABLES','BIRD','SOUP','CEREAL','CRACKERS','EGGS','CHEESE','FISH','CHICKEN','BEEF','BOWL','CUP','STEW',	20	CHICKEN	Related to [illegible]
'LETTER','ANIMAL','CAR','MAN','MAIL','FRIEND','DOG','STAMP','BOX','CARD','LICK','STICKER','SEND','DELIVER','SEAL','PACKAGE','MAILMAN', OFFICE','POSTAGE','PARCEL','WHOM',	18	ENVELOPE	Related to letter, mail, envelope and mail
'MUSIC','WORK','MONEY','COLLECT','CARDS','JOB','MESSAGE','PHONE','TALK','STATE','BILL','NOTE','TELEPHONE','CAPITAL','RECOMMENDATION',	17	LETTER	Related to alphabet, letters, ABC, numbe
'ALPHABET','MATH','ADD','NUMBERS','CALCULATOR','LOVE','LETTERS','ABC','KINDERGARTEN','TWENTY SIX','ARTS','STATISTICS','PLAY DOUGH','PAPERS',	14	NUMBER	Related to grade, work, class, school and
'MANY','NOTICE','TELEVISION','NEWSPAPER','NEWS','BOARD','BULLETIN','EDITORIAL','NUMEROUS','NETWORK','SERIES',	11	WORDS	Related to message, number, note and tel
'ONE','NUMBER','NAME','SIGN','FIRST','COUNT','SYMBOL','DIGIT','INITIAL','SERIAL',	10	JOB	Related to one, number, name, symbol an
'WORD','DICTIONARY','ENGLISH','VOCABULARY','WORDS','GREEK','FRATERNITY','SENTENCE','LANGUAGE','PARAGRAPH',	10	GIVE	Related to words, dictionary, word and en
'POST','OFFICE','COMPUTER','SECRETARY','TYPEWRITER','TYPIST',	6	OPENER	Related to notice, apper, newspaper and
'HEAD','ADDRESS','HOME','HOUSE','BLOCK',	5	BULLETIN	Non Relevant
'TAKE','GIVE','RECEIVE','GET',	4	SOUP	Related to A, B, letters,
'A','B','VITAMIN',	3	HOME	Non Relevant
'CAN','CLOSE','OPENER',	3	SECRETARY	Related to many
'CHOWDER','CLAM',	2	CHOWDER	Non Relevant

```
      "Related to letter, mail, envelope and mailbox",
      "Related to alphabet, letters, ABC, numbers and papers",
      "Related to  grade, work, class, school and money",
      "Related to message, number, note and telephone",
      "Related to one, number, name, symbol and digits",
      "Related to words, dictionary, word and english",
      "Related to notice, apper, newspaper and series",
      "Non Relevant",
      "Related to A, B, letters, ",
      "Non Relevant",
      "Related to many",
      "Non Relevant"
    )
  )

output_PG_2 %>%
  kbl(caption = "TABLE 18: INTERPRETATION FOR THE CLUSTER") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 4) %>%
  column_spec(column = 1, width = "5in")
```

```
#output_RW_1
View(output_PG_2)
```

Comments: Here, we can see the clusters, size of the cluster and one word label for the community. The community's label are stored in the central_label for each cluster. As we can see that the model on Page Rank is performing very good and also giving satisfactory results and outputs, as there are only two clusters which are non relevant to the target pairs and other all are related to target pairs. As the size of the cluster is reducing still the page rank is giving related words to the target pair.

```
output_PG_3 <- lables_community(PG_3, community_table(PG_3))
```

```
## [1] 10
## [1] 10
```

```
tibble(output_PG_3, .name_repair = "minimal") %>%
  kbl(caption= "TABLE 19: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER
      FOR PAGE RANK FOR THIRD PAIR OF TARGETS") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 6) %>%
  column_spec(column = 1, width = "5in")
```

Table 19: TABLE 19: SHOWING THE CENTRAL LABEL AND THE SIZE OF THE CLUSTER FOR PAGE RANK FOR THIRD PAIR OF TARGETS

C	S	central_label
'PAIN','CHILD','SCHOOL','KIDS','ABUSE','WIN','HAPPY','PLAY','YOUNG','FUN','LOVE','ADOLESCENT','KID','JUVENILE','BABY','CHILDREN','PARENTS DOUGH','TORTURE','MAZE',	28	TAKE
'AWAY','HERE','WITHOUT','GONE','LOST','FOOD','WATER','CAR','GO','LEAVE','PLACE','MAP','APPEAR','DISAPPEAR','COME','FIND','REMOVE','STOL	25	CHILD
'KIDNAP','STEAL','CAPTURE','MONEY','BAD','BOOK','ACQUIRE','WRONG','CRIME','LIE','CRIMINAL','JAIL','THIEF','ROBBER','CROOK','BURGLAR','C	24	TOGETHER
'TAKE','ACCEPT','GIVE','RECEIVE','GET','OBTAIN','HAVE','OUT','HOLD','KEEP','GRAB','CATCH','CARRY','BORROW','LOSE','LEND','BRING','ESCAPE'	22	HOW
'ANIMAL','ABSENCE','PRESENT','SICK','PRESENCE','TARDY','ATTENDANCE','ALONE','EMPTY','ABSENT','THERE','PEOPLE','HURT','PERSON','LATE'	18	TAKE
'ABDUCT','ADDUCT','MISSING','OPEN','TOGETHER','ACTION','WITH','VANISH','CONNECT','LINK','SAUSAGE','REAPPEAR','MILK CARTON',	13	GROWN-UPS
'GOOD','MINDED','GIFT','RIGHT','MOVIE','SCARY','HORRIBLE','LEFT','GIVING',	9	RIGHT
'ALIEN','ADULT','PARENT','OLD','ADULTS','GROWN-UPS','CITIZEN','PAST',	8	LOST
'STOP','MOVEMENT','RUN','WALK','ROAM',	5	RUN
'REACTION','CAT','DOG','BRACELET','CHAIN',	5	DOG

Table 20: TABLE 20: INTERPRETATION FOR THE CLUSTER

C	S	central_label	Explanation
'PAIN','CHILD','SCHOOL','KIDS','ABUSE','WIN','HAPPY','PLAY','YOUNG','FUN','LOVE','ADOLESCENT','KID','JUVENILE','BABY','CHILDREN','PARENTS DOUGH','TORTURE','MAZE',	28	TAKE	Related to pain and torchure
'AWAY','HERE','WITHOUT','GONE','LOST','FOOD','WATER','CAR','GO','LEAVE','PLACE','MAP','APPEAR','DISAPPEAR','COME','FIND','REMOVE','STOL	25	CHILD	Related to absent, disappear, stolen, hidden and lost
'KIDNAP','STEAL','CAPTURE','MONEY','BAD','BOOK','ACQUIRE','WRONG','CRIME','LIE','CRIMINAL','JAIL','THIEF','ROBBER','CROOK','BURGLAR','C	24	TOGETHER	Related to kidnap, steal, crime, jail, thief, robber, rob and robbery
'TAKE','ACCEPT','GIVE','RECEIVE','GET','OBTAIN','HAVE','OUT','HOLD','KEEP','GRAB','CATCH','CARRY','BORROW','LOSE','LEND','BRING','ESCAPE'	22	HOW	Related to snatch, lose and escape
'ANIMAL','ABSENCE','PRESENT','SICK','PRESENCE','TARDY','ATTENDANCE','ALONE','EMPTY','ABSENT','THERE','PEOPLE','HURT','PERSON','LATE'	18	TAKE	Related to not present, empt, absence, alone and late
'ABDUCT','ADDUCT','MISSING','OPEN','TOGETHER','ACTION','WITH','VANISH','CONNECT','LINK','SAUSAGE','REAPPEAR','MILK CARTON',	13	GROWN-UPS	Related to abduct, adduct, missing and vanish
'GOOD','MINDED','GIFT','RIGHT','MOVIE','SCARY','HORRIBLE','LEFT','GIVING',	9	RIGHT	Related to scarry and horrible
'ALIEN','ADULT','PARENT','OLD','ADULTS','GROWN-UPS','CITIZEN','PAST',	8	LOST	Non Relevant
'STOP','MOVEMENT','RUN','WALK','ROAM',	5	RUN	Non Relevant
'REACTION','CAT','DOG','BRACELET','CHAIN',	5	DOG	Non Relevant

```
output_PG_3 <- output_PG_3 %>%
  mutate(
    Explanation = c(
       "Related to pain and torchure",
       "Related to absent, disappear, stolen, hidden and lost",
       "Related to kidnap, steal, crime, jail, thief, robber, rob and robbery",
       "Related to snatch, lose and escape",
       "Related to not present, empt, absence, alone and late",
       "Related to abduct, adduct, missing and vanish ",
       "Related to scarry and horrible",
       "Non Relevant",
       "Non Relevant",
       "Non Relevant"
     )
   )

output_PG_3 %>%
  kbl(caption = "TABLE 20: INTERPRETATION FOR THE CLUSTER") %>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left", font_size = 4) %>%
  column_spec(column = 1, width = "5in")
```

```
#output_RW_1
View(output_PG_3)
```

Comments: Here, we can see the clusters, size of the cluster and one word label for the community. The community's label are stored in the central_label for each cluster. As we can see that the model on Page Rank is performing very good and also giving satisfactory results and outputs, as there very few clusters which are non relevant to the target pairs and other all are related to target pairs. As the size of the cluster is reducing still the page rank is giving related words to the target pair.

In conclusion to Page Rank performance on the respective target pairs, the community detection algorithms are working perfectly and returning just one to three non relevant clusters and other all clusters are relevant to the target pair of nodes. This model performed very well and gave satisfactory results as outcome for each cluster.

Finally concluding and comparing the working of random walk and page rank on community detection algorithm, Page rank performed very well and gave satisfactory results as compared to random walk in all the three pairs of target.