assignment3

Yong J-Hsen

Set up

```
rm(list=ls())
library(slam)
library(tm)
library(SnowballC)
library(dplyr)
library(ggplot2)
library(lsa)
library(igraph)
library(bipartite)
library(cluster)
set.seed(32025998)
```

Q1. Data collection

The broad topic that I chose was video games. I found fifteen articles talking about the good, the bad, and the evolution of gaming and its industry, 5 articles for each topic. All articles are cited in the references below.

Q2. Data processing

I selected parts of the articles that were relevant to their specific category. I compiled the text into .txt documents which I keep inside a folder called corpus, the folder has 15 .txt documents.

```
cname = file.path(".", "corpus")
docs = Corpus(DirSource((cname)))
print(summary(docs))
```

```
##           Length Class            Mode
## bad1.txt  2      PlainTextDocument list
## bad2.txt  2      PlainTextDocument list
## bad3.txt  2      PlainTextDocument list
## bad4.txt  2      PlainTextDocument list
## bad5.txt  2      PlainTextDocument list
## evo1.txt  2      PlainTextDocument list
## evo2.txt  2      PlainTextDocument list
## evo3.txt  2      PlainTextDocument list
## evo4.txt  2      PlainTextDocument list
## evo5.txt  2      PlainTextDocument list
## good1.txt 2      PlainTextDocument list
## good2.txt 2      PlainTextDocument list
## good3.txt 2      PlainTextDocument list
## good4.txt 2      PlainTextDocument list
## good5.txt 2      PlainTextDocument list
```

Q3. Text processing

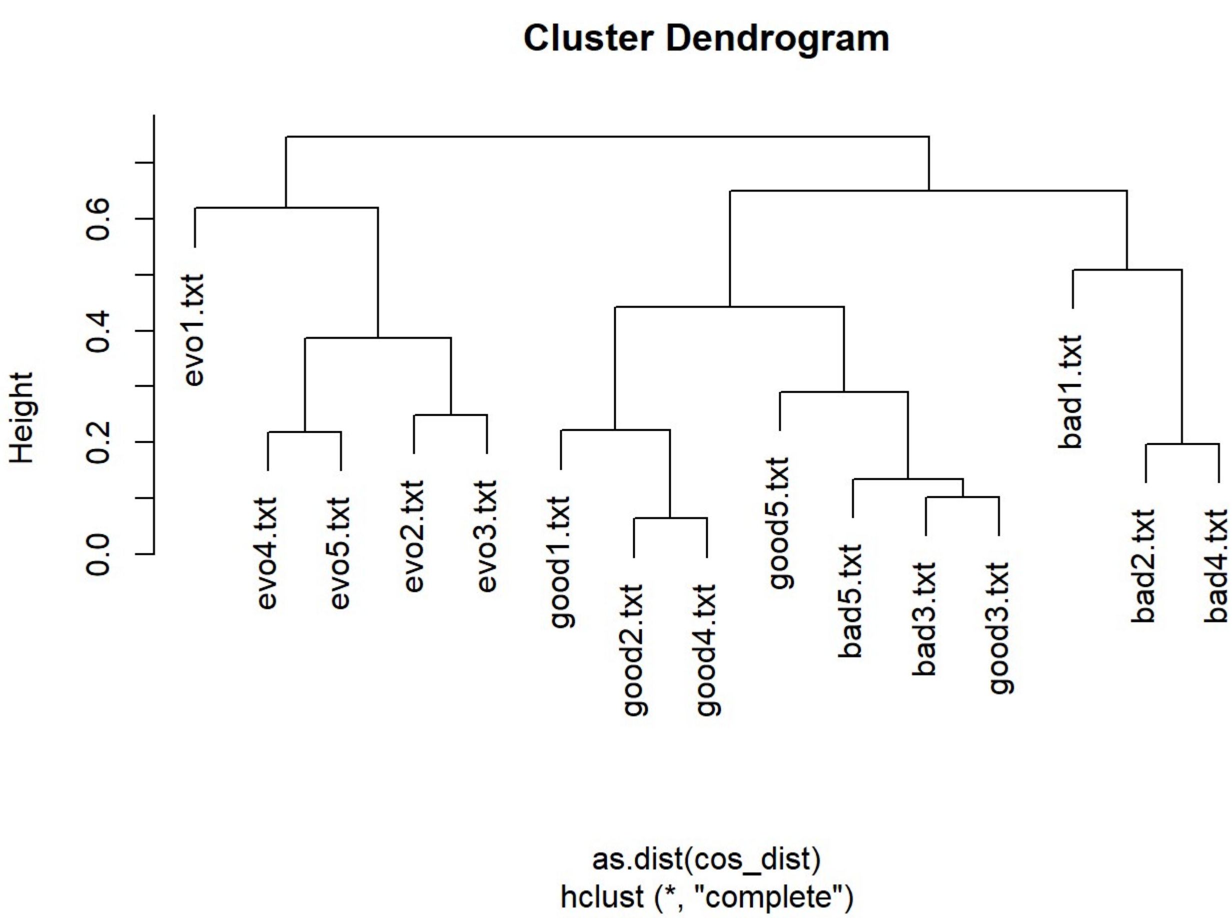
Applied standard text processing as taught in lecturers before creating the corpus such as removing numbers, white spaces, stop words, punctuation, etc. Also added “also”,“can”, and “game” to the stop words as they were the most commonly occurring words and removing them improving clustering results significantly. The finished dtm has 27 terms.

```
# Change all to lower case, remove numbers, punctuation, white spaces
docs <- tm_map(docs, content_transformer(tolower))
docs <- tm_map(docs, removeNumbers)
docs <- tm_map(docs, removePunctuation)
docs <- tm_map(docs, stripWhitespace)
# stemming, removes prefix/suffix of words
docs <- tm_map(docs, stemDocument, language = "english")
# Custom list of words to not include
custom_stopwords <- c(stopwords("english"), "also", "can","game")
docs <- tm_map(docs, removeWords, custom_stopwords)
#Create document term matrix
dtm <- DocumentTermMatrix(docs)
dtm = removeSparseTerms(dtm,sparse= 0.45)
#inspect(dtm)
dtmf = as.data.frame(as.matrix(dtm))
dtmf[1:10]
```

```
##           activ becom import includ lead like mani mental need one
## bad1.txt      1     2      1      2    1    4    1      5    1   2
## bad2.txt      6     2      1      2    6    0    4      3    2   1
## bad3.txt      3     2      3      3    0    2    2      9    2   3
## bad4.txt      3     1      0      1    5    0    1      1    1   2
## bad5.txt      3     0      1      5    2    2    5      5    3   1
## evo1.txt      0     0      0      4    0    2    1      0    0   2
## evo2.txt      0     3      1      0    0    1    0      0    0   0
## evo3.txt      0     1      0      0    0    2    0      0    0   1
## evo4.txt      0     1      0      1    1    0    0      0    0   0
## evo5.txt      0    10      1      0    0    3    2      0    2   1
## good1.txt     6     1      0      0    1    0    0      1    0   1
## good2.txt     4     0      7      2    6    2    3      2    4   2
## good3.txt     0     1      1      2    1    2    2     12    2   0
## good4.txt     3     0      0      0    1    2    3      2    0   2
## good5.txt     6     0      1      9    1    0    4      2    2  10
```

Q4. Hierarchical clustering

The clustering with 27 terms seems to perform well, with a classification accuracy of 86%. The evolution group is all correctly classified but there are some mistakes in the good/bad, could be due to the similarity in terms used. The silhouette score shows that most of the silhouette width is positive, which suggests that most nodes are correctly clustered, but there is still some values close to 0 and 1 negative number.



```
## The silhouette score:
```

```
##       cluster neighbor   sil_width
##  [1,]       1        2  0.03551693
##  [2,]       1        2  0.37117356
##  [3,]       2        3  0.50049173
##  [4,]       1        2  0.04124885
##  [5,]       2        1  0.40341509
##  [6,]       3        1  0.20117660
##  [7,]       3        2  0.07812200
##  [8,]       3        1  0.34839176
##  [9,]       3        2  0.10889894
## [10,]       3        2 -0.32587210
## [11,]       2        1  0.37942710
## [12,]       2        3  0.53408169
## [13,]       2        1  0.52282575
## [14,]       2        1  0.62007708
## [15,]       2        3  0.34288510
## attr(,"Ordered")
## [1] FALSE
## attr(,"call")
## silhouette.default(x = cluster_labels, dist = as.dist(cos_dist))
## attr(,"class")
## [1] "silhouette"
```

```
## The clustering matrix:
```

```
##            Clusters
## GroupNames  1 2 3
##   bad       3 2 0
##   evolution 0 0 5
##   good      0 5 0
```

```
## the Accuracy of the matrix is:  0.8666667
```

Q5. Single-mode network

Evo4 has the average score, making it the “most” important node. A transitivity of 1 means that the graph is transitive and that all nodes are able to reach all other nodes via other nodes. The degree score indicates that all nodes have the max number of edges and that all nodes have some similarity to other nodes.

The key node(evo4) has its edges colored red , the node have been colored based on their category for easier grouping, and the edge width thickness is based on the edge weight.

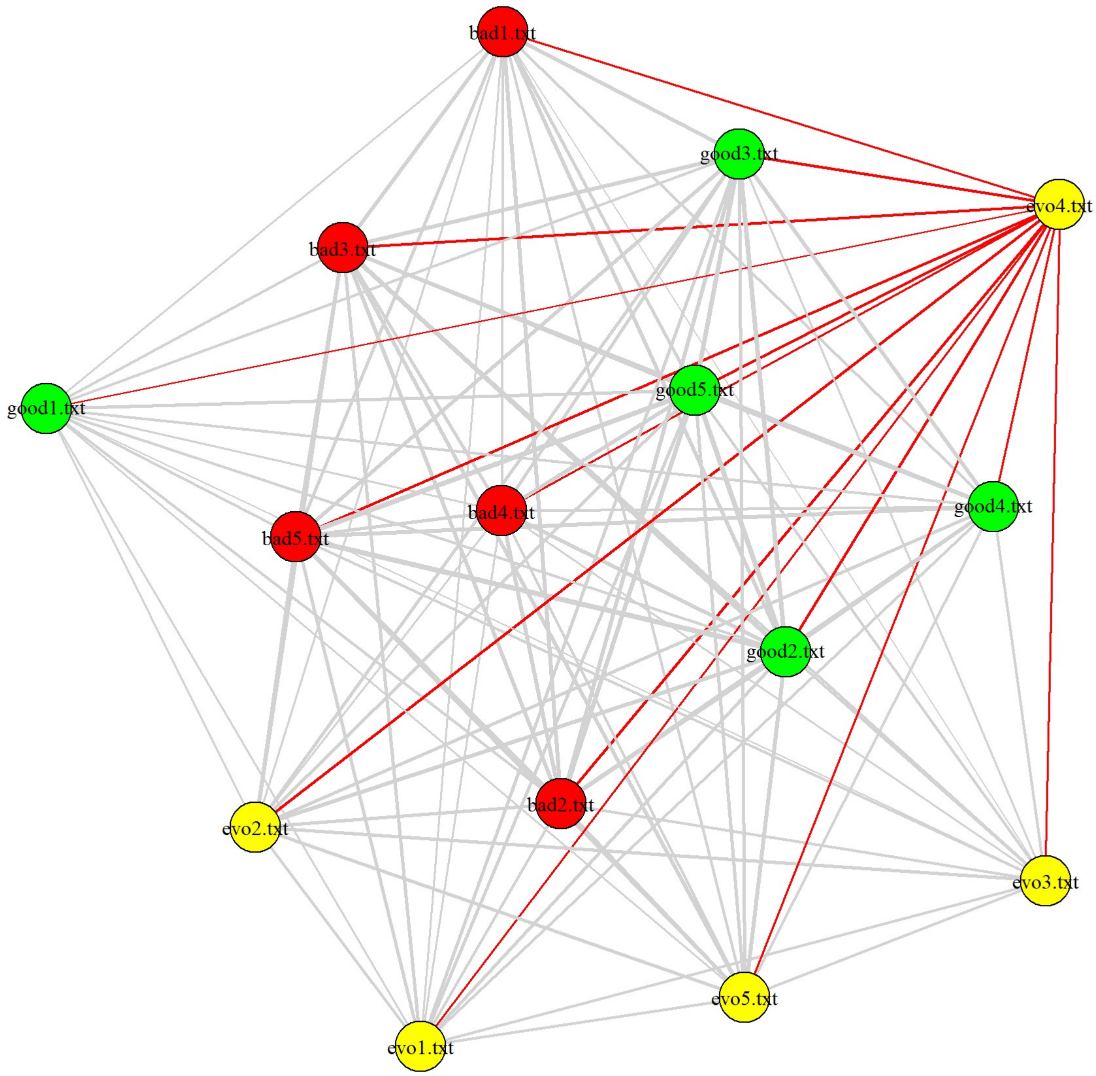
```
## The transitivity score:  1
```

```
## The table of network scores:
```

```
##           closeness degree betweenness Eigenvector      Avg
## bad1.txt     0.0060     14         0.0   0.7668727 4.668667
## bad2.txt     0.0047     14         0.0   0.9415762 4.668233
## bad3.txt     0.0050     14         0.0   0.9025413 4.668333
## bad4.txt     0.0060     14         0.0   0.7547783 4.668667
## bad5.txt     0.0047     14         0.0   0.9600728 4.668233
## evo1.txt     0.0062     14         0.0   0.7227163 4.668733
## evo2.txt     0.0058     14         0.0   0.7712919 4.668600
## evo3.txt     0.0072     14         0.0   0.6221752 4.669067
## evo4.txt     0.0078     14         3.5   0.5803902 5.835933
## evo5.txt     0.0058     14         0.0   0.7788294 4.668600
## good1.txt    0.0073     14         0.0   0.6286651 4.669100
## good2.txt    0.0045     14         0.0   1.0000000 4.668167
## good3.txt    0.0050     14         0.0   0.8961929 4.668333
## good4.txt    0.0054     14         0.0   0.8383158 4.668467
## good5.txt    0.0046     14         0.0   0.9719113 4.668200
```

```
## The most important row:
```

```
##          closeness degree betweenness Eigenvector      Avg
## evo4.txt    0.0078     14         3.5   0.5803902 5.835933
```



Q6. Single-mode network for words

The transitivity and degree of this network is similar to Q5, transitivity score of 1 and maximum number of edge connected for each node. “will” has the highest average score, making it the most important .

Similar to graph in Q5,the edge width is based off their weight, the key node is colored in yellow, and I avoided formatting the edges as it will make it very cluttered.

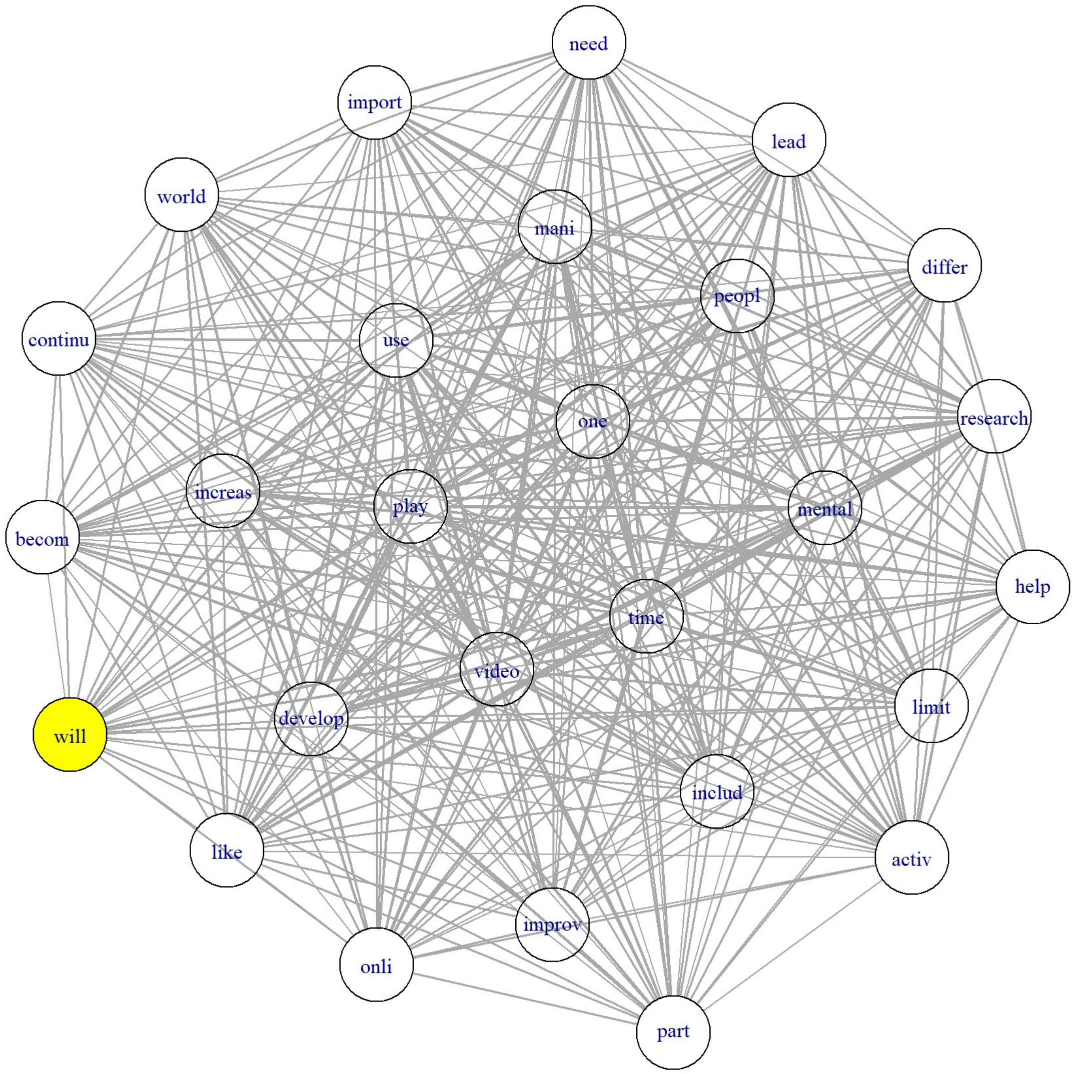
```
## The transitivity score:  1
```

```
## The table of network scores:
```

```
##          closeness degree betweenness Eigenvector      Avg
## activ       0.0059     26   0.0000000   0.6631394 8.668633
## becom       0.0061     26   0.3333333   0.6408955 8.779811
## import      0.0056     26   0.0000000   0.6843295 8.668533
## includ      0.0053     26   0.0000000   0.7212142 8.668433
## lead        0.0055     26   0.0000000   0.7067618 8.668500
## like        0.0055     26   0.0000000   0.6982068 8.668500
## mani        0.0048     26   0.0000000   0.8058873 8.668267
## mental      0.0053     26   0.0000000   0.7362821 8.668433
## need        0.0056     26   0.0000000   0.6826423 8.668533
## one         0.0046     26   0.0000000   0.8279721 8.668200
## peopl       0.0053     26   0.0000000   0.7318748 8.668433
## play        0.0040     26   0.0000000   0.9583609 8.668000
## research    0.0053     26   0.0000000   0.7206472 8.668433
## time        0.0042     26   0.0000000   0.8973256 8.668067
## use         0.0043     26   0.0000000   0.8943915 8.668100
## video       0.0038     26   0.0000000   1.0000000 8.667933
## world       0.0057     26   0.0000000   0.6727402 8.668567
## continu     0.0062     26   0.3333333   0.6212639 8.779844
## develop     0.0049     26   0.0000000   0.7834750 8.668300
## help        0.0058     26   0.0000000   0.6626762 8.668600
## increas     0.0051     26   0.0000000   0.7519124 8.668367
## onli        0.0060     26   0.0000000   0.6401480 8.668667
## part        0.0062     26   0.0000000   0.6201074 8.668733
## differ      0.0059     26   0.0000000   0.6564703 8.668633
## limit       0.0060     26   0.0000000   0.6494183 8.668667
## will        0.0062     26   3.5000000   0.6212766 9.835400
## improv      0.0060     26   0.0000000   0.6448560 8.668667
```

```
## The most important row:
```

```
##      closeness degree betweenness Eigenvector    Avg
## will    0.0062     26         3.5   0.6212766 9.8354
```



Q7. Bipartite graph

The bipartite graph does seem to show a discernible pattern. All evolution nodes are on the left side of the graph whereas bad and good can be seen on the right side as one big group. This could be due to similar words being used when describing the good and bad of video games without the context of the word taken into account.

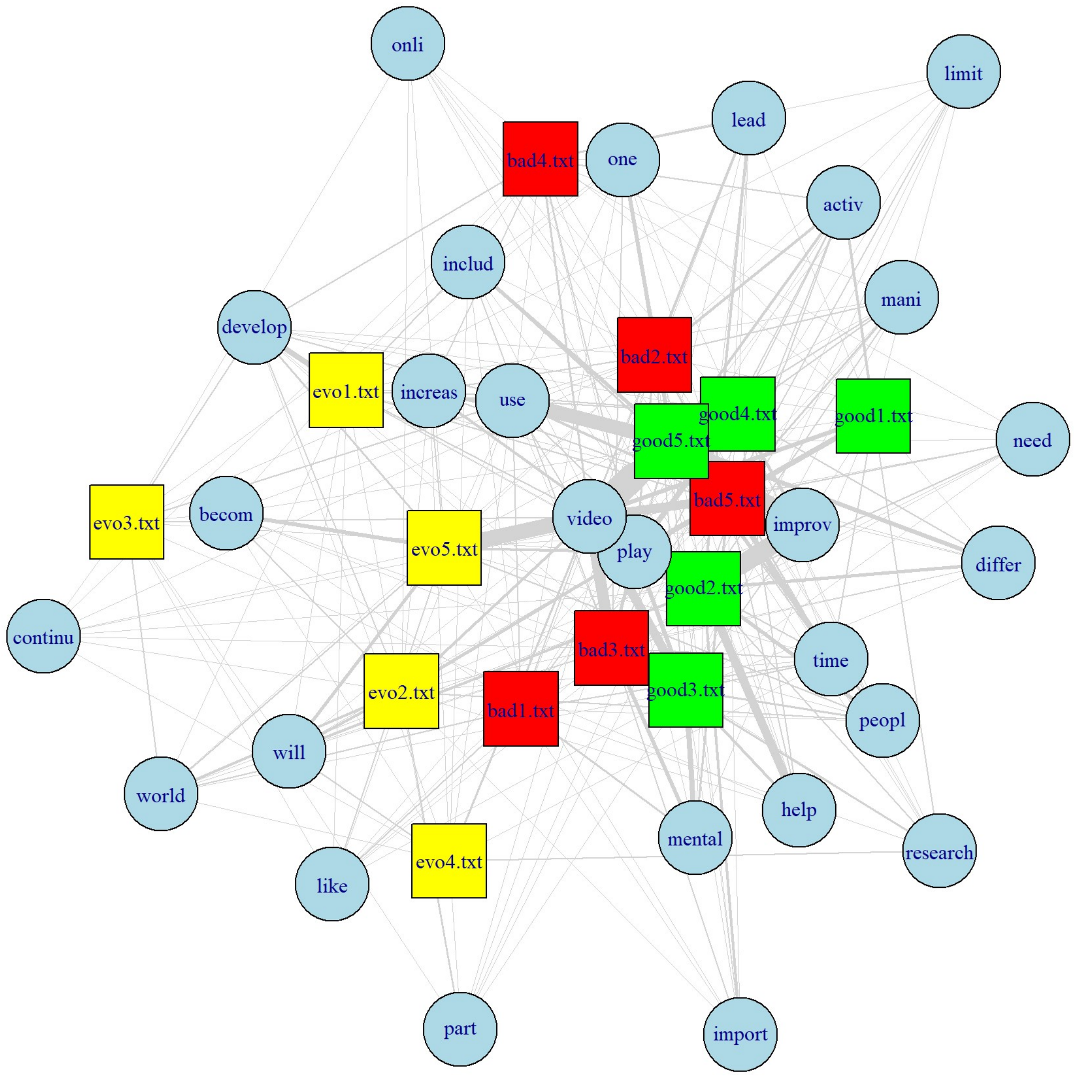
```
## The transitivity score:  0
```

```
## The table of network scores:
```

```
##            closeness degree betweenness Eigenvector       Avg
## bad1.txt      0.0105     17  67.6262848  0.05846091 28.212262
## bad2.txt      0.0116     22  63.9356075  0.07106818 28.649069
## bad3.txt      0.0108     21  38.4876486  0.26749613 19.832816
## bad4.txt      0.0109     17  66.8082157  0.06953477 27.939705
## bad5.txt      0.0106     23  34.0601042  0.25855342 19.023568
## evo1.txt      0.0109     16  52.4907664  0.06259129 22.833889
## evo2.txt      0.0104     18  54.8550004  0.11202698 24.288467
## evo3.txt      0.0103     14  39.8230233  0.04208029 17.944441
## evo4.txt      0.0109     13  55.2496807  0.04946389 22.753527
## evo5.txt      0.0102     18  12.5769374  0.34302384 10.195712
## good1.txt     0.0106     14  61.2237325  0.14663662 25.078111
## good2.txt     0.0093     24   5.7387987  0.82807056  9.916033
## good3.txt     0.0110     21  34.4624087  0.26035742 18.491136
## good4.txt     0.0099     19  20.3620434  0.24681451 13.123981
## good5.txt     0.0104     23  19.3107351  0.79144680 14.107045
## activ         0.0075      9   0.0000000  0.09296827  3.002500
## becom         0.0111     10  23.1123040  0.03988860 11.041135
## import        0.0111      9  46.1288764  0.06606500 18.379992
## includ        0.0093     10   1.6814034  0.09351821  3.896901
## lead          0.0111     10  34.7182613  0.06076726 14.909787
## like          0.0088     10   0.9565005  0.04127639  3.655100
## mani          0.0099     11  10.7554418  0.07700001  7.255114
## mental        0.0091     10   3.3194657  0.08772608  4.442855
## need          0.0093      9   2.1777760  0.05975558  3.729025
## one           0.0111     12  20.0189310  0.09418484 10.676677
## peopl         0.0101     10  13.0418935  0.13570880  7.683998
## play          0.0082     14   1.6723156  0.46360308  5.226839
## research      0.0100     10  11.7185114  0.05526869  7.242837
## time          0.0099     13  12.0321490  0.09299729  8.347350
## use           0.0101     13  11.0995935  0.25559616  8.036564
## video         0.0056     15   0.0000000  1.00000000  5.001867
## world         0.0101     10  13.2021763  0.03115818  7.737425
## continu       0.0109      9  20.2335961  0.02182441  9.748165
## develop       0.0104     12  17.4434425  0.05386754  9.817947
## help          0.0106      9  24.2585199  0.16523973 11.089707
## increas       0.0082     11   0.0000000  0.13168853  3.669400
## onli          0.0103      9  16.0063385  0.03982406  8.338880
## part          0.0105      9  19.1062798  0.02666903  9.372260
## differ        0.0106      9  18.5273785  0.13475413  9.179326
## limit         0.0115      9  55.4747966  0.02958230 21.495432
## will          0.0093      9   4.3250361  0.08468656  4.444779
## improv        0.0079      9   0.0000000  0.55648041  3.002633
```

```
## The most important row:
```

```
##           closeness degree betweenness Eigenvector      Avg
## bad2.txt    0.0116     22    63.93561  0.07106818 28.64907
```



Q8. Brief summary

When it comes to readability, the hierarchical clustering with a dendrogram is clearly more readable than the network and anyone can easily interpret the graph. However, if deeper analysis and insight is required, the network graph provides a more complex and comprehensive representation of the data as gives a detailed visualization of relationships, depicting both the direct and indirect interactions between nodes. Therefore, the best approach will depend on the situation.

References

1.Video gaming may be associated with better cognitive performance in children. (2022, October 24). National Institutes of Health (NIH). https://www.nih.gov/news-events/news-releases/video-gaming-may-be-associated-better-cognitive-performance-children (https://www.nih.gov/news-events/news-releases/video-gaming-may-be-associated-better-cognitive-performance-children)

2.Shatz, I. (2019). The Cognitive Benefits of Playing Video Games – Effectiviology. Effectiviology.com. https://effectiviology.com/cognitive-benefits-of-playing-video-games/ (https://effectiviology.com/cognitive-benefits-of-playing-video-games/)

3.Brennan, D. (2021, October 25). Are There Mental Health Benefits of Video Games? WebMD. https://www.webmd.com/mental-health/mental-health-benefits-of-video-games (https://www.webmd.com/mental-health/mental-health-benefits-of-video-games)

4.Group, T. G. (n.d.). The Benefits of Video Games & VR for Seniors – The Goodman Group. Blog.thegoodmangroup.com. https://blog.thegoodmangroup.com/best-video-games-for-seniors (https://blog.thegoodmangroup.com/best-video-games-for-seniors)

5.Jiménez-Muñoz, L., Peñuelas-Calvo, I., Calvo-Rivera, P., Díaz-Oliván, I., Moreno, M., Baca-García, E., & Porras-Segovia, A. (2021). Video Games for the Treatment of Autism Spectrum Disorder: A Systematic Review. Journal of Autism and Developmental Disorders, 52(1). https://doi.org/10.1007/s10803-021-04934-9 (https://doi.org/10.1007/s10803-021-04934-9)

6.Video Games and Mental Health: Risk Factors and Benefits. (2022, June 15). Psych Central. https://psychcentral.com/depression/can-video-games-cause-depression#how-gaming-can-cause-depression (https://psychcentral.com/depression/can-video-games-cause-depression#how-gaming-can-cause-depression)

7.Raymer, D. (2021, September 1). Internet gaming disorder: A real mental health issue on the rise in adolescents and young adults. Counseling Today. https://ct.counseling.org/2021/09/internet-gaming-disorder-a-real-mental-health-issue-on-the-rise-in-adolescents-and-young-adults/ (https://ct.counseling.org/2021/09/internet-gaming-disorder-a-real-mental-health-issue-on-the-rise-in-adolescents-and-young-adults/)

8.Video Games, Mental Health, and Addiction – The Good, The Bad, and The Ugly: Family Addiction Specialist: Addiction Counselor. (n.d.). Www.familyaddictionspecialist.com. https://www.familyaddictionspecialist.com/blog/video-games-mental-health-and-addiction-the-good-the-bad-and-the-ugly (https://www.familyaddictionspecialist.com/blog/video-games-mental-health-and-addiction-the-good-the-bad-and-the-ugly)

9.Grinspoon, P. (2020, December 22). The health effects of too much gaming. Harvard Health Blog; Harvard Health Publishing. https://www.health.harvard.edu/blog/the-health-effects-of-too-much-gaming-2020122221645 (https://www.health.harvard.edu/blog/the-health-effects-of-too-much-gaming-2020122221645)

10.Barden, A. (2022, June 2). Can Video Games Cause Health Problems? All about Vision. https://www.allaboutvision.com/resources/human-interest/video-games-and-your-health/ (https://www.allaboutvision.com/resources/human-interest/video-games-and-your-health/)

11.Quigley, K. (2022, December 15). Features of Unreal Engine 5 That Will Revolutionise the Gaming Industry. Starloop Studios. https://starloopstudios.com/features-of-unreal-engine-5-that-will-revolutionise-the-gaming-industry/ (https://starloopstudios.com/features-of-unreal-engine-5-that-will-revolutionise-the-gaming-industry/)

12.Abdulla, O. (2021, September 20). How gaming industry has evolved over the last 20 years. Https://Techxmedia.com/ (Https://Techxmedia.com/). https://techxmedia.com/how-gaming-industry-has-evolved-over-the-last-20-years/ (https://techxmedia.com/how-gaming-industry-has-evolved-over-the-last-20-years/)

13.The Evolution of the Gaming Industry. (2022, June 17). Https://Www.outlookindia.com/ (Https://Www.outlookindia.com/). https://www.outlookindia.com/business-spotlight/the-evolution-of-the-gaming-industry-news-202966 (https://www.outlookindia.com/business-spotlight/the-evolution-of-the-gaming-industry-news-202966)

14.Robson, K. (2023, April 3). Gaming industry to be worth $470bn by 2030 despite setbacks, experts predict. Verdict. https://www.verdict.co.uk/gaming-industry-to-be-worth-470bn-by-2030-despite-setbacks-experts-predict/ (https://www.verdict.co.uk/gaming-industry-to-be-worth-470bn-by-2030-despite-setbacks-experts-predict/)

15.Beattie, A. (2021, October 31). How the Video Game Industry Is Changing. Investopedia. https://www.investopedia.com/articles/investing/053115/how-video-game-industry-changing.asp (https://www.investopedia.com/articles/investing/053115/how-video-game-industry-changing.asp)