

Trabajo Final BAIN

Sergio Esteban Tarrero

21/5/2021

```
setwd("C:/Users/Sergio/Desktop/Trabajo_Final_BAIN")  
getwd()  
## [1] "C:/Users/Sergio/Desktop/Trabajo_Final_BAIN"  
load("enron_data_revised.rda")
```

Cargo las librerías que voy a utilizar

```
##  
## Attaching package: 'igraph'  
  
## The following objects are masked from 'package:stats':  
##  
##   decompose, spectrum  
  
## The following object is masked from 'package:base':  
##  
##   union  
  
##  
## Attaching package: 'gplots'  
  
## The following object is masked from 'package:stats':  
##  
##   lowess  
  
## Package version: 3.0.0  
## Unicode version: 13.0  
## ICU version: 69.1  
  
## Parallel computing: 8 of 8 threads used.  
  
## See https://quanteda.io for tutorials and examples.  
  
##  
## Attaching package: 'quanteda.textplots'  
  
## The following object is masked from 'package:igraph':  
##  
##   as.igraph
```

```
##
## Attaching package: 'wordcloud'

## The following object is masked from 'package:gplots':
##
##     textplot

##
## Attaching package: 'SentimentAnalysis'

## The following object is masked from 'package:base':
##
##     write
```

Parte 1 - SNA - Social Network Analysis

Creación del grafo

```
grafo_trabajo <- graph.data.frame(edges,
                                   directed = TRUE,
                                   vertices = nodes)
```

Información del grafo creado

```
class(grafo_trabajo)

## [1] "igraph"

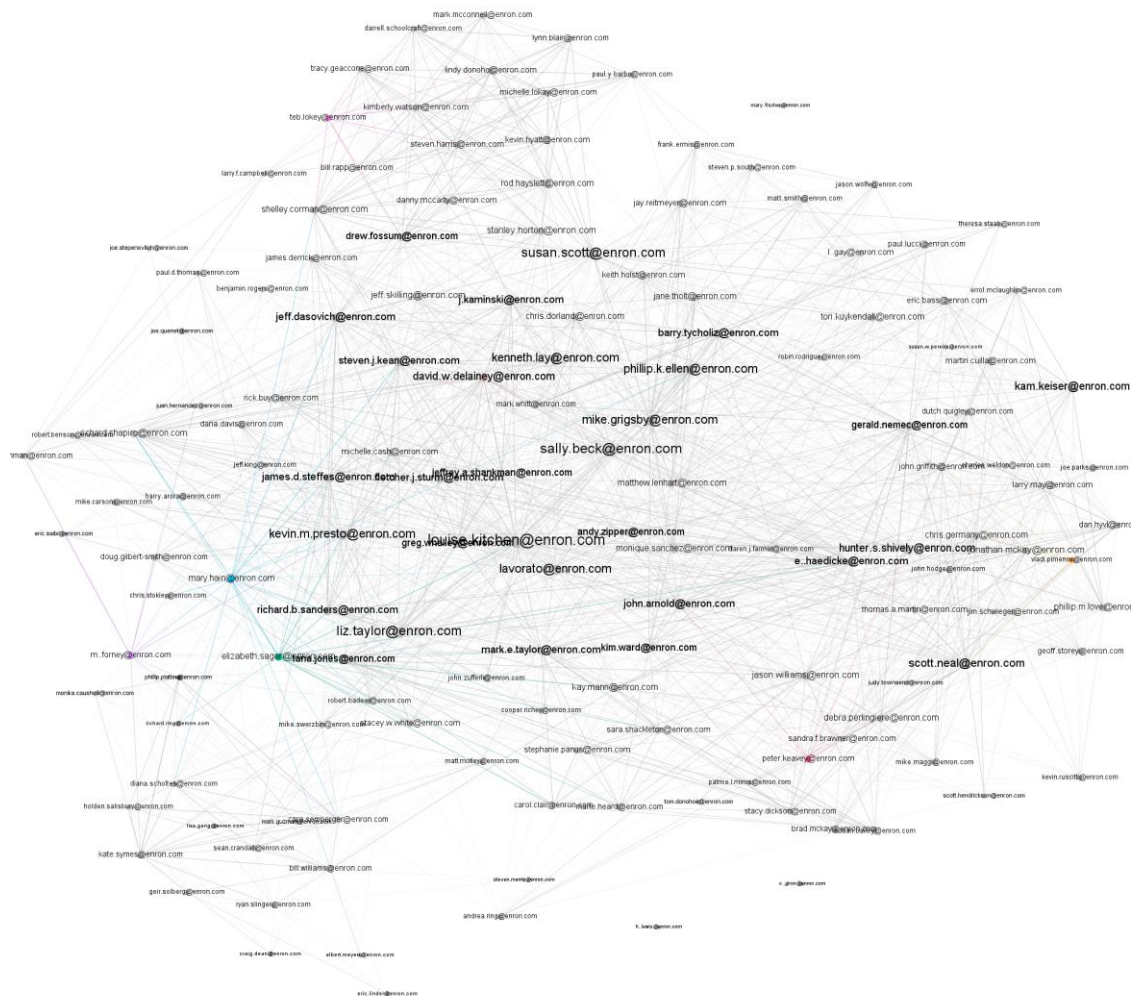
summary(grafo_trabajo)

## IGRAPH 1fcd8d9 DN-- 149 4308 --
## + attr: name (v/c), lastName (v/c), status (v/c), type (e/c), date
## | (e/n), count (e/n)
```

Exportar el grafo creado

```
write.graph(grafo_trabajo,
            file = "Grafo_Trabajo_BAIN_Sergio_Esteban_Tarrero.graphml",
            format = "graphml")
```

Grafo pasado por Gephi:



Grafo_Gephi

Cálculo de métricas individuales

Cálculo grado

2.1 Cálculo Grado (total, dado que es un grafo no dirigido)

```
nodes$degree_total <- degree(grafo_trabajo, v = V(grafo_trabajo), mode = c("total"))
```

```
nodes$degree_total
```

```
## [1] 59 117 84 13 61 59 88 70 74 48 96 73 13 233 143 129
90 24
## [19] 47 11 19 83 23 57 95 91 40 29 60 65 19 16 34 30
```

```

93 118
## [37] 54 68 33 90 53 114 23 89 7 23 0 79 100 22 46 56
35 59
## [55] 55 32 83 62 18 20 9 67 43 51 28 56 0 136 75 49
40 51
## [73] 110 47 91 16 55 5 44 77 34 21 40 101 87 30 59 125
47 15
## [91] 36 0 83 42 27 71 62 76 149 33 28 39 13 50 21 48
48 23
## [109] 41 114 19 61 45 10 113 97 69 42 25 22 86 23 122 47
80 20
## [127] 129 39 105 57 64 86 67 124 91 2 29 9 149 54 107 54
28 50
## [145] 15 57 58 101 22

```

Cálculo Betweenness

```
nodes$betweenness <- betweenness(grafo_trabajo, v=V(grafo_trabajo))
```

```
nodes$betweenness
```

```

## [1] 19.5955340 308.9832901 50.3889029 5.1306212 53.6508038
## [6] 16.8379893 784.4653611 491.2882056 608.1262083 73.4900115
## [11] 574.2400885 166.0263024 18.9056460 1565.5598693 1277.8826060
## [16] 697.1294257 140.3590779 78.0653882 87.7036414 5.8431749
## [21] 16.1579429 317.6467992 43.7692414 695.5040502 318.7819894
## [26] 139.2735655 64.4530830 29.2833549 117.8601898 169.7222848
## [31] 15.3658966 0.9045297 29.9666052 41.0446129 426.1825904
## [36] 656.0271828 43.1585998 196.4560413 10.2358912 180.8907777
## [41] 42.0679758 531.3648328 14.3542119 109.3271657 6.0802777
## [46] 30.9839019 0.0000000 140.9853022 335.8099856 117.2890481
## [51] 56.1481722 736.6559643 14.6683832 518.5893965 9.9736910
## [56] 57.4496077 558.3272059 262.1187011 12.2420145 85.2840706
## [61] 2.5003263 128.3758104 98.4434604 49.3506720 34.0472996
## [66] 25.1343831 0.0000000 369.5488268 172.1334972 152.8523062
## [71] 265.6378595 130.9392816 363.9773651 96.0691998 215.5110628
## [76] 8.2352217 119.0217509 0.0000000 46.8989777 268.9112471
## [81] 9.1300087 35.8993484 44.6816208 538.4336716 112.4993560
## [86] 25.3489082 63.4063767 369.9443594 45.3976408 6.0312109
## [91] 45.6789658 0.0000000 738.7324293 25.1051925 14.0552976
## [96] 117.4777468 36.4843646 55.3846048 964.5229778 13.2258477
## [101] 19.3015215 176.7842561 34.1488287 51.8455136 10.6229611
## [106] 39.5593661 65.7987159 25.8463997 50.2950290 459.5305517
## [111] 31.7739174 103.2434816 77.6628458 7.2689034 343.5460238
## [116] 116.1520439 30.3246233 162.3162756 14.3075656 4.9704852
## [121] 291.7610685 53.8757753 566.8058401 66.5963480 71.0013774
## [126] 9.4151726 722.7362519 121.4798686 396.2699942 254.8607115
## [131] 33.2782785 217.8969180 98.7498000 221.6350321 97.2343176
## [136] 0.0000000 10.0794443 6.2117080 2016.3283474 16.4413352

```

```
## [141] 340.5119055 28.3410733 17.7202713 60.1629275 8.1054242
## [146] 78.6057926 7.6817365 393.1751735 8.6330350
```

Cálculo Reach 2 step

```
nodes$reach_2_step <- neighborhood.size(grafo_trabajo, order = 2,
                                         nodes = V(grafo_trabajo), mode =
c("all"))
```

```
nodes$reach_2_step
```

```
## [1] 124 138 96 30 140 91 135 86 132 139 144 126 116 142 146 145
75 106
## [19] 131 102 109 126 127 141 139 144 126 97 121 126 126 117 129 93
137 142
## [37] 135 136 133 129 131 136 130 136 89 125 1 136 141 118 74 117
120 63
## [55] 108 126 130 130 122 125 92 106 140 128 123 82 1 142 110 125
136 137
## [73] 118 127 139 135 128 24 117 139 127 95 126 134 140 131 134 141
76 40
## [91] 120 1 140 118 134 131 138 96 140 133 131 137 85 130 123 125
89 129
## [109] 121 142 77 118 134 73 138 140 141 135 132 120 138 96 143 125
121 115
## [127] 138 122 126 132 110 136 125 142 97 23 131 109 140 100 139 80
119 128
## [145] 120 127 62 139 121
```

Cálculo Transitivity

```
transitivity(grafo_trabajo, type="global")
```

```
## [1] 0.4212257
```

```
transitivity(as.undirected(grafo_trabajo, mode="collapse"))
```

```
## [1] 0.4212257
```

```
transitivity(grafo_trabajo, type="local")
```

```
## [1] 0.06545880 0.03698792 0.03557085 0.11538462 0.14699454
0.06779661
## [7] 0.03996865 0.02815735 0.04776009 0.14893617 0.12258772
0.05251142
## [13] 0.60256410 0.02305017 0.04471585 0.07860950 0.03071161
0.07971014
## [19] 0.12210916 0.29090909 0.22807018 0.09168381 0.13043478
0.06077694
## [25] 0.07614782 0.14627595 0.10512821 0.07142857 0.06892655
```

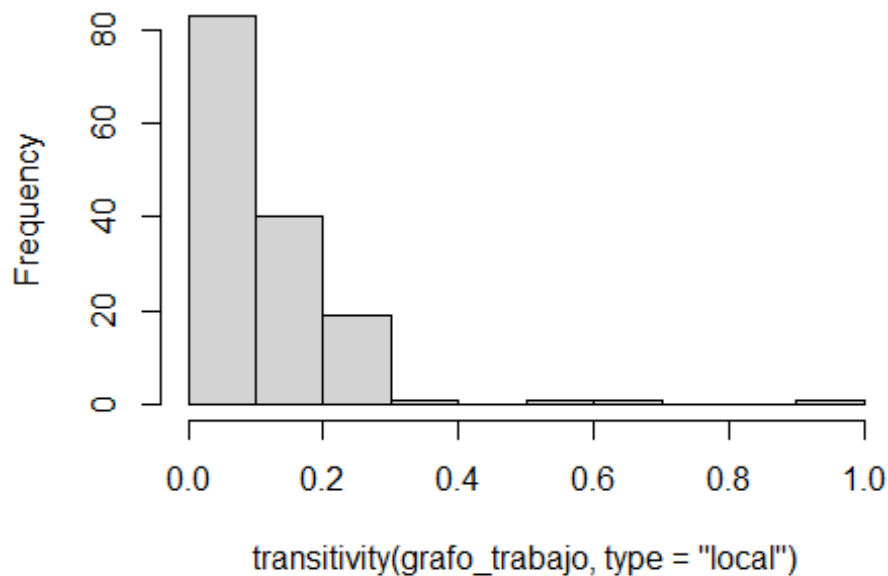
```

0.10432692
## [31] 0.19883041 0.25833333 0.16577540 0.07586207 0.08648901
0.03708533
## [37] 0.08525507 0.07374890 0.19507576 0.05243446 0.13207547
0.04021115
## [43] 0.22529644 0.06767109 0.23809524 0.24901186      NaN
0.10451152
## [49] 0.06464646 0.24242424 0.07439614 0.04415584 0.19831933
0.03857393
## [55] 0.06060606 0.12096774 0.03408757 0.08249603 0.18300654
0.22105263
## [61] 0.52777778 0.03934871 0.17165006 0.10039216 0.08730159
0.05519481
## [67]      NaN 0.04346405 0.04900901 0.05357143 0.12179487
0.09098039
## [73] 0.03352794 0.08233117 0.05421245 0.20833333 0.07003367
0.30000000
## [79] 0.06236786 0.12200957 0.21746881 0.15714286 0.17179487
0.03643564
## [85] 0.08393478 0.24367816 0.05493863 0.04167742 0.07308048
0.26666667
## [91] 0.21111111      NaN 0.04672348 0.08130081 0.15384615
0.07766600
## [97] 0.07456372 0.04350877 0.03854526 0.14962121 0.16402116
0.12145749
## [103] 0.28205128 0.20244898 0.17142857 0.09929078 0.06737589
0.19762846
## [109] 0.18658537 0.07607514 0.16374269 0.09234973 0.15151515
0.06666667
## [115] 0.04171934 0.04746564 0.08141517 0.10452962 0.29333333
0.16883117
## [121] 0.04076607 0.11067194 0.09497358 0.12765957 0.04841772
0.15263158
## [127] 0.05026647 0.06207827 0.02710623 0.07644110 0.05505952
0.06456908
## [133] 0.05382180 0.03763441 0.03296703 1.00000000 0.31280788
0.27777778
## [139] 0.03283149 0.05870021 0.04496561 0.07547170 0.08465608
0.12489796
## [145] 0.19047619 0.11842105 0.05989111 0.05683168 0.29437229

hist(transitivity(grafo_trabajo, type="local"))

```

Histogram of transitivity(grafo_trabajo, type = "local")



Vemos la data que hay

```
ls()
## [1] "edges"          "edges.full"     "grafo_trabajo" "network"
## [5] "nodes"
```

Añado una columna al grafo que vamos a generar con los datos obtenidos de la

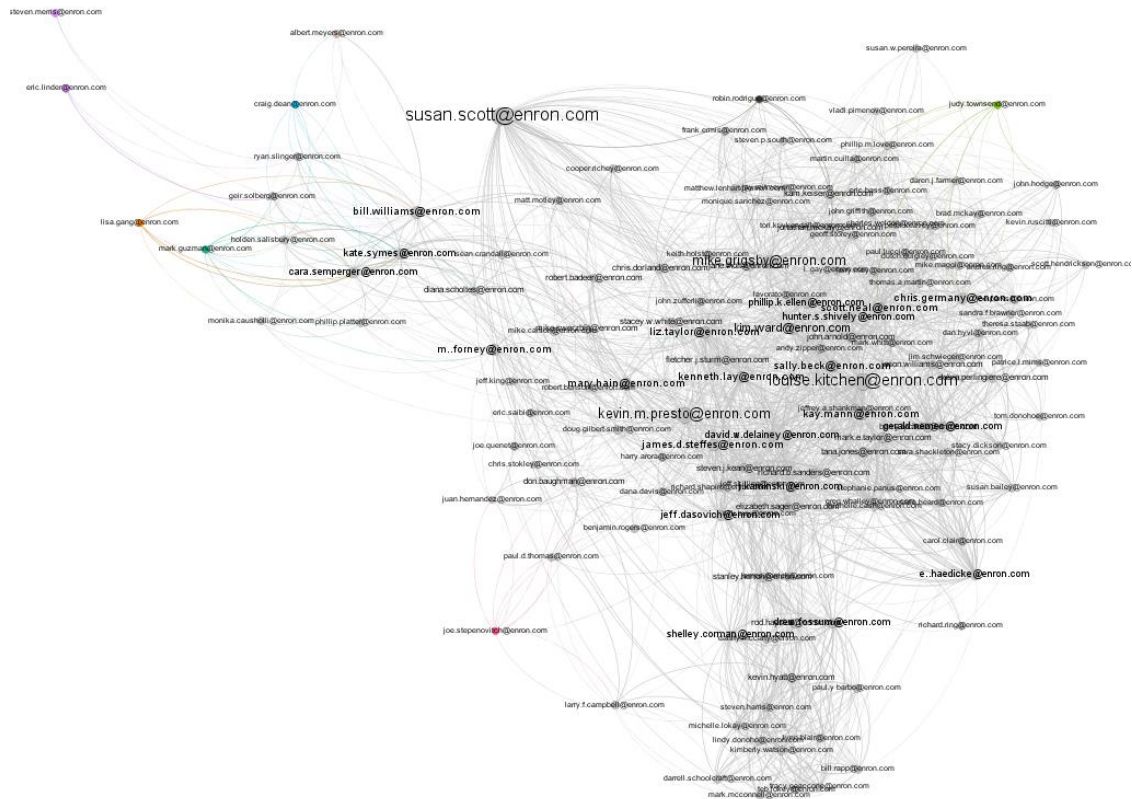
Betweenness, que hemos calculado antes

```
nodes$betweenness <- betweenness(grafo_trabajo, v=V(grafo_trabajo))
V(grafo_trabajo)$betweenness<-nodes$betweenness
```

Exporta el grafo con la columna Betweenness añadida

```
write.graph(grafo_trabajo, file="Grafo_Gephi_Con_Betweenness.graphml",
format="graphml")
```

Grafo organizado por comunidades:



Grafo_Gephi_Con_Betweenness

Parte 2 - TM - Text Mining

En este apartado vamos a ver de que han hablado los trabajadores de Enron,

más específico, los CEOS de la empresa

ls()


```
## [1] "edges"          "edges.full"      "grafo_trabajo"  "network"
## [5] "nodes"

summary(nodes)

##      Email_id      lastName      status
degree_total
## Length:149      Length:149      Length:149      Min.    :
0.00
## Class :character Class :character Class :character 1st Qu.:
28.00
## Mode  :character Mode  :character Mode  :character Median :
53.00
##                                     Mean    :
57.83
##                                     3rd Qu.:
83.00
##                                     Max.    :
233.00
##      betweenness      reach_2_step
## Min.    : 0.0      Min.    : 1.0
## 1st Qu.: 19.6      1st Qu.:110.0
## Median : 65.8      Median :126.0
## Mean    :188.1      Mean    :117.5
## 3rd Qu.:221.6      3rd Qu.:136.0
## Max.    :2016.3     Max.    :146.0

summary(edges.full)

##      sender      receiver      type      subject
## Length:61673     Length:61673     Length:61673     Length:61673
## Class :character Class :character Class :character Class
:character
## Mode  :character Mode  :character Mode  :character Mode
:character
##      body      date
## Length:61673     Length:61673
## Class :character Class :character
## Mode  :character Mode  :character
```

CEOS de Enron, Información de los CEOS, Matriz con datos

Para sacar los mensajes necesito el correo electrónico "Email_id"

```
ceos_con_info <- nodes[nodes$status == "CEO",]
ceos_enron <- nodes[nodes$status == "CEO", "Email_id"]

ceos_con_info
```

	Email_id	lastName	status	degree_total	betweenness
## 5	jeff.skilling@enron.com	Skilling	CEO	61	53.6508
## 11	kenneth.lay@enron.com	Lay	CEO	96	574.2401
## 26	lavorato@enron.com	Lavorato	CEO	91	139.2736
## 68	david.w.delainey@enron.com	Delainey	CEO	136	369.5488

	reach_2_step
## 5	140
## 11	144
## 26	144
## 68	142

Voy a trabajar con “ceos_enron” porque el “ceos_con_info” es para dar más datos

de los CEOS (Incluye el grado, betweenness y reach 2 step)

Mensajes Enviados

```
mensajes_enviados <- edges.full$body[edges.full$sender %in% ceos_enron]
```

Cantidad de mensajes mandados

```
length(mensajes_enviados)
```

```
## [1] 1381
```

Muestra algunos mensajes completos mandados por los CEOS

```
head(mensajes_enviados)
```

```
## [1] "The Stanford Business School, in collaboration with the
management-consulting firm McKinsey & Company, is conducting a large
study of global corporations, their management practices, and
organization. Enron is participating in the project in order to learn
how our own practices and activities compare to those of other leading
global firms and to see which bring true benefits in the global
marketplace. One of the unique features of the project is that it seeks
to assess the knowledge and communication networks within a company by
asking employees throughout the company to comment on their actual
practices and experience. The goal is to obtain real and accurate
information by surveying the people most directly involved in the company
s operations. I would like to ask you to help us in getting the most out
of our participation in the GLOBE Initiative by completing the GLOBE
Network Survey. The survey is brief and should take no longer than 15 or
20 minutes. Your responses will be treated as strictly confidential, and
will be used solely by the Stanford/McKinsey researchers to evaluate how
specific practices and structures impact global communication and
knowledge flows. If you would like individual feedback on how your own
```

networks compare to those of others in the study, please check the appropriate box at the end of the survey. Please save this email for future reference and, again, thank you for contributing to this important learning opportunity. To complete the GLOBE Network Survey, please click on the link below and simply follow the survey instructions. You will need to enter the Enron Company Code, which is: C673EV To begin the survey, click on the link below:
<http://63.211.208.100/GLOBENetwork/survey.asp> You may also copy this link and paste it into your favorite browser. Thank you."

Mensajes Recibidos

```
mensajes_recibidos <- edges.full$body[edges.full$receiver %in%  
ceos_enron]
```

Cantidad de mensajes recibidos

```
length(mensajes_recibidos)
```

```
## [1] 913
```

Muestra algunos mensajes completos recibidos por los CEOs

```
head(mensajes_recibidos)
```

```
## [1] "Dear Dr. Lay--\t\tThank you for your presentation yesterday. As  
always, you handled very well some tough questions in a very trying time  
for Enron. I was literally mortified when I heard the \"crack\"  
question-- but I think you handled that well too. As a 17-year Enron  
veteran, the question made me truly ashamed to be associated with the co-  
worker who asked the question. Hindsight is always 20/20. I m sure you  
must know that the majority of Enron s employees don t feel the same way  
as this one individual. Thanks for your continued hard work and  
support. Kevin Hyatt Director, Asset DevelopmentETS"
```

Juntar los mensajes recibidos y mandados

```
todos_mensajes <- c(mensajes_recibidos, mensajes_enviados)
```

```
# Número total de mensajes recibidos y mandados  
length(todos_mensajes)
```

```
## [1] 2294
```

Limpieza de texto

Poner todo el texto en letra minúscula

```
todos_mensajes <- tolower(todos_mensajes)
```

```
head(todos_mensajes)
```

```
## [1] "dear dr. lay--\t\tthank you for your presentation yesterday. as  
always, you handled very well some tough questions in a very trying time  
for enron. i was literally mortified when i heard the \"crack\  
question-- but i think you handled that well too. as a 17-year enron  
veteran, the question made me truly ashamed to be associated with the co-  
worker who asked the question. hindsight is always 20/20. i m sure you  
must know that the majority of enron s employees don t feel the same way  
as this one individual. thanks for your continued hard work and  
support.kevin hyattdirector, asset developmentets"
```

Limpieza de caracteres

```
todos_mensajes <- str_replace_all(todos_mensajes, pattern = "[:;,]", " ")  
todos_mensajes <- str_replace_all(todos_mensajes, pattern = "[-+*/]", "  
")  
todos_mensajes <- str_replace_all(todos_mensajes, pattern = "[?!\\"]", "  
")  
todos_mensajes <- str_replace_all(todos_mensajes, pattern = "\\\\\\", " ")  
todos_mensajes <- str_replace_all(todos_mensajes, pattern = "\\t", " ")  
todos_mensajes <- str_replace_all(todos_mensajes, pattern = "&", " ")  
todos_mensajes <- str_replace_all(todos_mensajes, pattern = " ", " ")  
todos_mensajes <- str_replace_all(todos_mensajes, pattern = " ", " ")
```

```
head(todos_mensajes)
```

```
## [1] "dear dr. lay  thank you for your presentation yesterday. as  
always you handled very well some tough questions in a very trying time  
for enron. i was literally mortified when i heard the crack question  but  
i think you handled that well too. as a 17 year enron veteran the  
question made me truly ashamed to be associated with the co worker who  
asked the question. hindsight is always 20 20. i m sure you must know  
that the majority of enron s employees don t feel the same way as this  
one individual. thanks for your continued hard work and support.kevin  
hyattdirector asset developmentets"
```

Creación del corpus utilizando la librería Quanteda

```
corpus_ceos <- corpus(todos_mensajes)
```

```
summary(corpus_ceos)
```

```
## Corpus consisting of 2294 documents, showing 100 documents:
```

```
##
```

```
##      Text Types Tokens Sentences
```

```
##      text1      79      112         1
```

```
##      text2       3         3         1
```

```
##      text3       3         3         1
```

```
##      text4     101      185         1
```

```
##      text5     101      185         1
```

```
##      text6      21       24         1
```

```
##      text7      91      141         1
```

```
##      text8      91      141         1
```

```
##      text9      82      122         1
```

```
##      text10    103      149         1
```

```
##      text11    103      149         1
```

```
##      text12     66      104         1
```

```
##      text13    311      687         2
```

```
##      text14     18       21         1
```

```
##      text15    103      203         1
```

```
##      text16     46       56         1
```

```
##      text17     46       56         1
```

```
##      text18     75      101         2
```

```
##      text19     75      101         2
```

```
##      text20    716     1862         1
```

```
##      text21    574     1580         1
```

```
##      text22    118      173         1
```

```
##      text23     47       88         1
```

```
##      text24    228      485         1
```

```
##      text25    232      552         1
```

```
##      text26    232      552         1
```

```
##      text27    150      280         1
```

```
##      text28    150      280         1
```

```
##      text29    316      736         1
```

```
##      text30    292      611         1
```

```
##      text31    293      597         1
```

```
##      text32     73      110         1
```

```
##      text33    103      170         2
```

```
##      text34    103      170         2
```

```
##      text35    321      652         2
```

```
##      text36    112      160         1
```

```
##      text37     61       77         1
```

```
##      text38    163      304         2
```

```
##      text39    422     1108         1
```

```
##      text40    199      479         1
```

```
##      text41    174      260         1
```

```
##      text42    289      661         1
```

##	text43	20	20	1
##	text44	215	448	1
##	text45	261	522	1
##	text46	149	284	1
##	text47	217	423	2
##	text48	440	1017	1
##	text49	72	105	1
##	text50	28	30	1
##	text51	79	114	1
##	text52	142	225	1
##	text53	123	188	1
##	text54	122	210	1
##	text55	34	37	1
##	text56	69	100	1
##	text57	48	58	1
##	text58	8	8	1
##	text59	10	11	1
##	text60	19	22	1
##	text61	50	63	1
##	text62	98	167	1
##	text63	142	272	1
##	text64	241	398	1
##	text65	241	398	1
##	text66	149	253	1
##	text67	90	132	1
##	text68	67	99	1
##	text69	63	78	1
##	text70	69	88	1
##	text71	377	727	1
##	text72	416	820	1
##	text73	123	209	1
##	text74	428	1100	1
##	text75	428	1100	1
##	text76	176	380	1
##	text77	176	380	1
##	text78	94	137	1
##	text79	61	77	1
##	text80	29	39	1
##	text81	27	33	1
##	text82	27	33	1
##	text83	61	93	1
##	text84	61	93	1
##	text85	61	93	1
##	text86	61	93	1
##	text87	131	207	1
##	text88	184	278	1
##	text89	184	278	1
##	text90	43	52	1
##	text91	43	52	1
##	text92	70	96	1

```
## text93 70 96 1
## text94 256 502 1
## text95 137 207 1
## text96 81 119 1
## text97 25 25 1
## text98 50 64 1
## text99 153 246 1
## text100 222 334 1
```

Eliminar cosas innecesarias (urls, numeros sueltos, signos de puntuación)

```
palabras <- tokens(corpus_ceos,
  remove_punct = TRUE,
  remove_numbers = TRUE,
  remove_url = TRUE)

class(palabras)

## [1] "tokens"
```

Palabras más repetidas en los correos

```
myStemMat <- dfm(palabras)

myStemMat <- dfm_remove(myStemMat, stopwords("english"))

topfeatures(myStemMat, 100)
```

##	=	enron	hou
s			
##	4604	4392	3888
3187			
##	program	analyst	associate
>			
##	3060	2616	1974
1898			
##	one	also	business
company			
##	1840	1513	1496
1473			
##	please	new	<
david			
##	1389	1339	1321
1315			
##	can	\$	corp
subject			
##	1163	1138	1135
1133			

##	jeff	continue	power
na			
##	1022	994	975
948			
##	w	many	may
t			
##	920	915	913
880			
##	know	group	message
delainey			
##	870	861	857
848			
##	market	california	bill
important			
##	832	816	798
794			
##	year	committee	said
prc			
##	785	780	773
743			
##	j	edison	remains
john			
##	736	724	723
720			
##	purpose	attached	created
leaders			
##	720	707	695
694			
##	direction	e	time
people			
##	691	684	683
680			
##	mark	current	graduates
committeeassociate			
##	679	648	640
640			
##	date	august	energy
ect			
##	631	625	624
623			
##	forwarded	meeting	us
provide			
##	617	609	591
591			
##	like	issue	office
get			
##	580	555	539
539			
##	d	memo	pm
sent			

##	538	535	534
530			
##	state	original	gas
need			
##	530	527	521
506			
##	last	ken	ena
davis			
##	506	502	501
498			
##	lay	pmt	back
first			
##	497	495	494
494			
##	key	making	peer
development			
##	493	487	481
478			
##	better	friday	puc
plan			
##	472	466	465
460			
##	future	leading	every
comments			
##	457	457	451
448			
##	now	end	groups
however			
##	447	444	444
443			

```
mystopwords <- c(stopwords("english"),
  "=", "s", ">", "<",
  "e", "j", "d", "w", "t", "$")
```

```
myStemMat <- dfm_remove(myStemMat, mystopwords)
```

```
topfeatures(myStemMat, 100)
```

##	enron	hou	program
analyst			
##	4392	3888	3060
2616			
##	associate	one	also
business			
##	1974	1840	1513
1496			
##	company	please	new
david			
##	1473	1389	1339

1315			
##	can	corp	subject
jeff			
##	1163	1135	1133
1022			
##	continue	power	na
many			
##	994	975	948
915			
##	may	know	group
message			
##	913	870	861
857			
##	delainey	market	california
bill			
##	848	832	816
798			
##	important	year	committee
said			
##	794	785	780
773			
##	prc	edison	remains
john			
##	743	724	723
720			
##	purpose	attached	created
leaders			
##	720	707	695
694			
##	direction	time	people
mark			
##	691	683	680
679			
##	current	graduates	committeeassociate
date			
##	648	640	640
631			
##	august	energy	ect
forwarded			
##	625	624	623
617			
##	meeting	us	provide
like			
##	609	591	591
580			
##	issue	office	get
memo			
##	555	539	539
535			
##	pm	sent	state

original			
##	534	530	530
527			
##	gas	need	last
ken			
##	521	506	506
502			
##	ena	davis	lay
pmto			
##	501	498	497
495			
##	back	first	key
making			
##	494	494	493
487			
##	peer	development	better
friday			
##	481	478	472
466			
##	puc	plan	future
leading			
##	465	460	457
457			
##	every	comments	now
end			
##	451	448	447
444			
##	groups	however	lavorato
support			
##	444	443	440
439			
##	call	james	opportunities
dwr			
##	435	433	431
427			
##	resources	times	customers
questions			
##	426	424	422
421			

Generación de Bigramas y Trigramas

```
palabras2 <- tokens_select(palabras, pattern = mystopwords, selection = "remove")
```

```
palabras3 <- tokens_ngrams(palabras2, n = 2:3)
```

```
matriz2 <- dfm(palabras3)
```

Las 100 palabras más repetidas

```
topfeatures(matriz2, 100)
```

```
##          associate_analyst          analyst_program
##                1961                1283
##  associate_analyst_program          david_delainey
##                1281                645
##                program_one committeeassociate_analyst
##                640                640
##                delainey_hou          david_delainey_hou
##                611                610
##                hou_ect          original_message
##                591                409
##                jeff_skilling          office_chairman
##                401                350
##                embedded_picture          ken_lay
##                344                341
##                enron_company          billy_lemmons
##                334                329
##                one_can          picture_metafile
##                325                325
##                one_enron          new_businesses
##                324                323
##                making_enron embedded_picture_metafile
##                323                323
##                company_also          exposure_enron
##                322                322
##                world_leading          subject_associate
##                322                321
##                program_program          program_many
##                321                321
##                analyst_committee subject_associate_analyst
##                321                321
##                program_worldwide          worldwide_ken
##                320                320
##                lay_department          department_office
##                320                320
##                chairman_subject          program_date
##                320                320
##                date_august          august_know
##                320                320
##                know_jeff          skilling_created
##                320                320
##                created_associate          one_successful
##                320                320
##                successful_important          important_contributions
##                320                320
##                contributions_enron          enron_continue
##                320                320
##                continue_cornerstone          cornerstone_company
```

##	320	320
##	company_remains	remains_single
##	320	320
##	single_important	important_source
##	320	320
##	source_new	new_talent.the
##	320	320
##	talent.the_purpose	purpose_memo
##	320	320
##	memo_reassure	reassure_direction
##	320	320
##	direction_philosophy	philosophy_importance
##	320	320
##	importance_program	program_remains
##	320	320
##	remains_unchanged	unchanged_said
##	320	320
##	said_many	many_times
##	320	320
##	times_new	businesses_created
##	320	320
##	created_graduates	graduates_members
##	320	320
##	members_program	many_current
##	320	320
##	current_leaders	leaders_company
##	320	320
##	also_graduates	graduates_program.three
##	320	320
##	program.three_broad	broad_principles
##	320	320
##	principles_continue	continue_guide
##	320	320
##	guide_direction	direction_program
##	320	320
##	one_associate	program_enron
##	320	320
##	company_asset	asset_one
##	320	320
##	enron_presence	presence_campus
##	320	320
##	campus_provide	provide_clarity
##	320	320
##	clarity_message	message_purpose
##	320	320
##	purpose_program	program_rotational
##	320	320
##	rotational_nature	nature_prepare
##	320	320
##	prepare_future	future_leaders

```
##          320          320
##      leaders_success      success_providing
##          320          320
```

Wordcloud

```
set.seed(100)
png(filename = "Wordcloud_Enron_Trabajo_Final_BAIN.png",
     width = 5000,
     height = 5000)
textplot_wordcloud(matriz2,
                   min_count = 20,
                   random_order = FALSE,
                   rotation = 0,
                   color = RColorBrewer::brewer.pal(8, "Dark2"))
```

Workcloud creada:



Wordcloud_Enron

Tópicos:

```
time1 <- Sys.time()

quant_dfm <- dfm_trim(matriz2,
                      min_termfreq = 20)

set.seed(100)
if (require(topicmodels)) {
  my_lda_fit12 <- LDA(convert(quant_dfm, to = "topicmodels"),
                     k = 6)
  get_terms(my_lda_fit12, 5)
}

##      Topic 1      Topic 2      Topic 3      Topic 4
## [1,] "peer_group" "look_forward" "david_delainey"
## [2,] "plan_attend"
## [2,] "hou_ect"      "please_call"      "delainey_hou"
## [2,] "morgan_stanley"
## [3,] "year_end"      "take_place"      "david_delainey_hou"
## [3,] "enron_entitled"
## [4,] "business_units" "monday_october"  "hou_ect"
## [4,] "new_business"
## [5,] "david_oxley"    "questions_please" "forwarded_david"
## [5,] "business_school"

##      Topic 5      Topic 6
## [1,] "associate_analyst" "issue_comes"
## [2,] "analyst_program"   "original_message"
## [3,] "associate_analyst_program" "direct_access"
## [4,] "program_one"        "please_mail"
## [5,] "committeeassociate_analyst" "hello_everyone.aep"

time2 <- Sys.time()
print(time2-time1)

## Time difference of 22.33297 secs

qq <- my_lda_fit12@beta

class(qq)

## [1] "matrix" "array"

dim(qq)

## [1] 6 2959

colnames(qq) <- my_lda_fit12@terms

qq[, 5:10]
```



```
##      interview_session session_quality quality_may original_message
## [1,]      -205.207525      -235.93447  -237.19067      -48.712081
## [2,]      -178.515392      -214.25629  -213.91449      -5.878840
## [3,]      -196.694105      -228.78857  -229.77698      -5.326552
## [4,]      -196.671690      -229.37006  -230.39569      -5.726184
## [5,]      -213.709791      -244.80620  -244.57362     -68.314730
## [6,]        -6.092971        -6.47246   -6.47246      -4.501530
##      hello_everyone.aep may_resemble
## [1,]      -209.995614      -236.90070
## [2,]      -189.194114      -211.60350
## [3,]      -207.695760      -227.93408
## [4,]      -207.507420      -229.56452
## [5,]      -224.689561      -244.23700
## [6,]        -5.012058        -6.47246
```

Visualización de texto

```
png(file = "Visualizacion_Texto.png",
    width = 5000,
    height = 5000,
    res = 300,
    bg = "black")

par(mfrow=c(6, 1))

for (k in 1:length(qq[,1])) {

  topic1 <- qq[k,]

  v <- topic1

  # utilizando rank pasamos el beta numérico a orden (entero, positivo)
  d <- data.frame(word = names(v), rank= rank(v))

  # ordenamos descendente (por defecto -sin el "-" es ascendente)
  d <- d[order(-d$rank),]

  # normalizamos (parecido a una frecuencia de palabras) +100 para que
  # tenga rango amplio
  d$freq <- d$rank - max(d$rank) + 100

  # Now with a prettier layout
  # baed on code published in
  # http://onertipaday.blogspot.com.es/2011/07/word-cloud-in-r.html
  #plot.new()

  pal2 <- brewer.pal(11, "Spectral")
  wordcloud(d$word,
            d$freq,
```

