Formula 1 Overtakes Analysis

Alessandro Relato

2025-05-21

Formula 1 Overtakes Analysis

Introduction

With this project I aim to study and analyze the overtakes happened in Formula 1 from 1994 to 2020. I will do two types of analysis: one general analysis with a single graph that includes all the overtakes together and one year per year with 27 graphs and compare the results to see how overtakes has changed during the years. The overtakes in this analysis include only the overtakes done on track (no overtakes done thanks to the strategy during the pit stops). The overtakes during the start have not been calculated because the start is always a mess and the overtakes happened thanks to an error made by one driver are not calculated too (only the unforced errors).

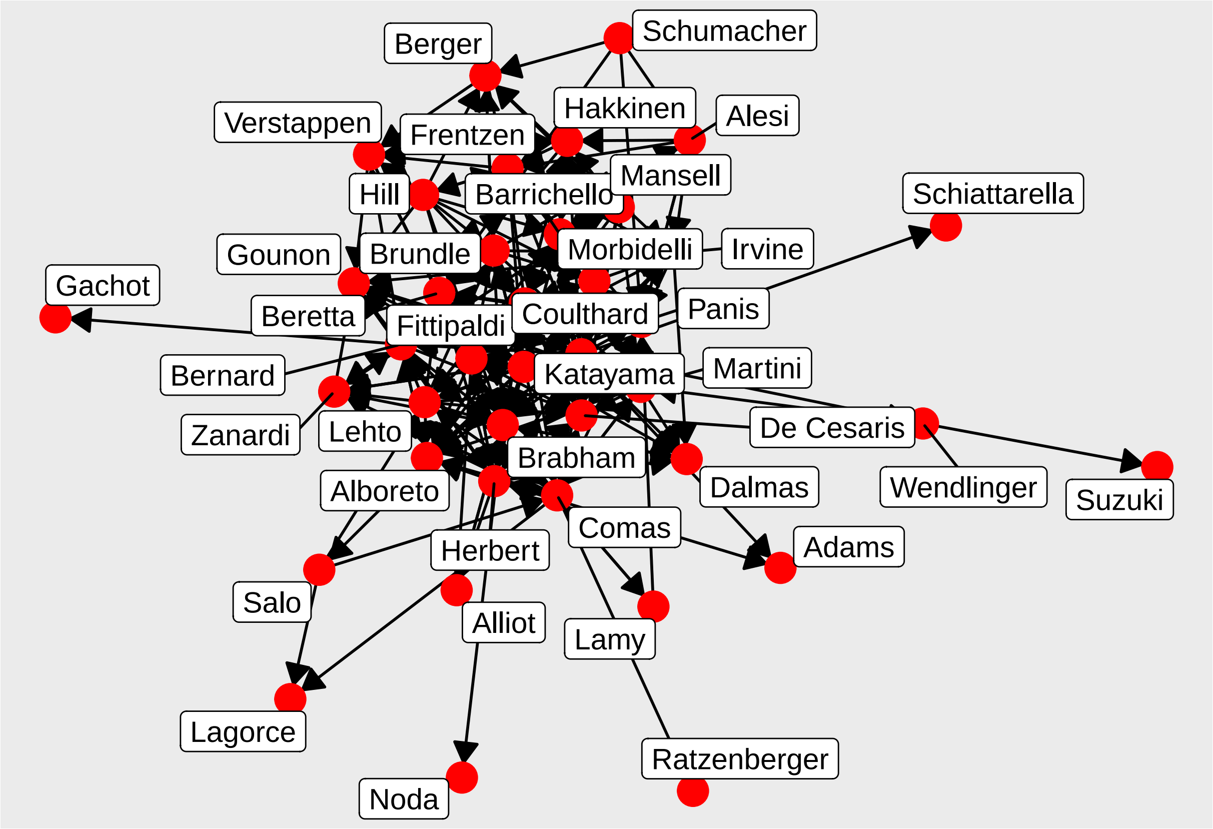
The dataset

The general graph is a weighted and directed graph and has X nodes and Y edges. The annual graphs are weighted and directed as well and have less nodes, about twenty each one. All these graph are created from a dataset containing all the information about overtakes on those years. I extract only the ones that I needed (name of overtaker, name of overtakee and year) and then I turned them into nodes and edges.

```r
library(dplyr)
library(ggplot2)
library(igraph)
library(tidygraph)
library(ggraph)
library(tibble)
library(tidyr)
```

```r
# Importing and creating general graph
nodes_general <- read.csv("nodes.csv", encoding = "UTF-8")
edges_general <- read.csv("edges.csv", encoding = "UTF-8")
g <- graph_from_data_frame(d = edges_general, vertices = nodes_general, directed = TRUE)

# Importing and creating annual graphs
g_yearly <- list()
for(year in 1994:2020) {
  nodes_year <- read.csv(paste0("yearly/", year, " nodes.csv"), encoding = "UTF-8")
  edges_year <- read.csv(paste0("yearly/", year, " edges.csv"), encoding = "UTF-8")
  g_yearly[[as.character(year)]] <- graph_from_data_frame(d = edges_year,
                                                          vertices = nodes_year, directed = TRUE)
}

# Example graph (plot of data from 1994)
ggraph(g_yearly[["1994"]], layout = "graphopt", niter = 5000) +
    geom_edge_link(end_cap = circle(2, "mm"), start_cap = circle(2, "mm"),
                   arrow = arrow(type = "closed", length = unit(3, "mm"))) +
    geom_node_point(color = "red", size = 5) +
    geom_node_label(aes(label = name), repel = TRUE)
```

```
## Warning: ggrepel: 1 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



General infos

```
# Calculating general data
infos <- matrix(NA, nrow = length(g_yearly), ncol = 4)
colnames(infos) <- c("Nodes", "Edges", "Density", "Reciprocity")
rownames(infos) <- names(g_yearly)
for(year in 1994:2020) {
  infos[as.character(year), "Nodes"] <- vcount(g_yearly[[as.character(year)]])
  infos[as.character(year), "Edges"] <- ecount(g_yearly[[as.character(year)]])
  infos[as.character(year), "Density"] <- edge_density(g_yearly[[as.character(year)]])
  infos[as.character(year), "Reciprocity"] <- reciprocity(g_yearly[[as.character(year)]])
}

variance <- c(var(infos[, "Nodes"]), var(infos[, "Edges"]), var(infos[, "Density"]),
              var(infos[, "Reciprocity"]))
media <- c(mean(infos[, "Nodes"]), mean(infos[, "Edges"]), mean(infos[, "Density"]),
           mean(infos[, "Reciprocity"]))
infos <- rbind(infos, media)
infos <- rbind(infos, variance)
infos <- rbind(infos, c(vcount(g), ecount(g), edge_density(g), reciprocity(g)))
rownames(infos) <- c(names(g_yearly), "Mean", "Variance", "General")
infos
```

```
##               Nodes     Edges    Density Reciprocity
## 1994       40.00000  218.0000 0.13974359  0.29357798
## 1995       35.00000  249.0000 0.20924370  0.39357430
## 1996       24.00000  153.0000 0.27717391  0.33986928
## 1997       26.00000  238.0000 0.36615385  0.47899160
## 1998       23.00000  162.0000 0.32015810  0.24691358
## 1999       24.00000  187.0000 0.33876812  0.28877005
## 2000       23.00000  224.0000 0.44268775  0.50892857
## 2001       26.00000  192.0000 0.29538462  0.39583333
## 2002       23.00000  186.0000 0.36758893  0.41935484
## 2003       24.00000  219.0000 0.39673913  0.42922374
## 2004       25.00000  245.0000 0.40833333  0.45714286
## 2005       25.00000  177.0000 0.29500000  0.37288136
## 2006       27.00000  225.0000 0.32051282  0.40000000
## 2007       26.00000  263.0000 0.40461538  0.57034221
## 2008       22.00000  225.0000 0.48701299  0.54222222
## 2009       25.00000  208.0000 0.34666667  0.52173913
## 2010       27.00000  323.0000 0.46011396  0.40866873
## 2011       29.00000  391.0000 0.48152709  0.58312020
## 2012       25.00000  404.0000 0.67333333  0.62376238
## 2013       23.00000  331.0000 0.65415020  0.62839879
## 2014       23.00000  320.0000 0.63241107  0.55625000
## 2015       21.00000  263.0000 0.62619048  0.53992395
## 2016       25.00000  344.0000 0.57333333  0.52906977
## 2017       24.00000  244.0000 0.44202899  0.42622951
## 2018       20.00000  253.0000 0.66578947  0.59288538
## 2019       20.00000  259.0000 0.68157895  0.61776062
## 2020       23.00000  260.0000 0.51383399  0.56153846
## Mean       25.11111  250.4815 0.43778051  0.47136936
## Variance   17.48718 4212.9516 0.02242524  0.01186853
## General   161.00000 3493.0000 0.13559783  0.64891180
```

In this table I summarized the general info for all the graph and for the general one. I also add the mean and the variance calculated on the annual graphs. As we can see the variance of reciprocity and density is low, so even if the number of nodes and edges are different there is a common structure. Obviously in the general graph the density is lower because we have much more nodes and the edges are only between drivers that has competed in the same years. On the other hand the reciprocity is higher because some drivers may overtook another one in one season but never in the previous or in the following season. As we can see the density values are not high, in fact the mean is 0.4. I was expecting that because the data does not include lapping, so the top drivers have not a lot of connections with the drivers on the back of the grid. Talking about reciprocity, instead, I was expecting to find it a little bit higher because usually there are some groups of drivers that overtake each other during a season (top drivers, mid-field, guys in the back).

Centrality analysis

Degree centrality

```r
# Degree
cen_degree <- matrix(NA, nrow = length(g_yearly), ncol = 10)
colnames(cen_degree) <- c("Max in", "Max in Name", "Min in", "Min in Name",
                          "Mean in", "Max out", "Max out Name", "Min out", "Min out Name", "Mean out")
rownames(cen_degree) <- names(g_yearly)
for(year in 1994:2020) {
  cen_degree[as.character(year), "Max in"] <- max(degree(g_yearly[[as.character(year)]],
                                                         mode = "in"))
  cen_degree[as.character(year), "Max out"] <- max(degree(g_yearly[[as.character(year)]],
                                                          mode = "out"))
  cen_degree[as.character(year), "Min in"] <- min(degree(g_yearly[[as.character(year)]],
                                                         mode = "in"))
  cen_degree[as.character(year), "Min out"] <- min(degree(g_yearly[[as.character(year)]],
                                                          mode = "out"))
  cen_degree[as.character(year), "Mean in"] <- mean(degree(g_yearly[[as.character(year)]],
                                                           mode = "in"))
  cen_degree[as.character(year), "Mean out"] <- mean(degree(g_yearly[[as.character(year)]],
                                                            mode = "out"))
  cen_degree[as.character(year), "Min in Name"] <- names(sort(
                                                    degree(g_yearly[[as.character(year)]],
                                                           mode = "in")))[[1]]
  cen_degree[as.character(year), "Max in Name"] <- names(sort(
                                                    degree(g_yearly[[as.character(year)]],
                                                           mode = "in"), decreasing = TRUE))[[1]]
  cen_degree[as.character(year), "Min out Name"] <- names(sort(
                                                    degree(g_yearly[[as.character(year)]],
                                                           mode = "out")))[[1]]
  cen_degree[as.character(year), "Max out Name"] <- names(sort(
                                                    degree(g_yearly[[as.character(year)]],
                                                           mode = "out"), decreasing = TRUE))[[1]]
}
cen_degree <- rbind(cen_degree, c(max(degree(g, mode = "in")), names(sort(
                                                                  degree(g, mode = "in"),
                                                                  decreasing = TRUE))[[1]],
                                  min(degree(g, mode = "in")), names(sort(
                                                                  degree(g, mode = "in")))[[1]],
                                  mean(degree(g, mode = "in")),
                                  max(degree(g, mode = "out")), names(sort(
                                                                  degree(g,
                                                                      mode = "out"),
                                                                      decreasing = TRUE))[[1]],
                                  min(degree(g, mode = "out")), names(sort(
                                                                  degree(g,
                                                                      mode = "out")))[[1]],
                                  mean(degree(g, mode = "out"))))
```

```r
rownames(cen_degree) <- c(names(g_yearly), "General")
cen_degree
```

```
##         Max in Max in Name      Min in Min in Name    Mean in            Max out
## 1994    "17"   "Brabham"     "0"    "Schumacher"  "5.45"             "21"   
## 1995    "15"   "Inoue"       "0"    "Tarquini"    "7.11428571428571" "16"   
## 1996    "12"   "Lamy"        "1"    "Marques"     "6.375"            "15"   
## 1997    "17"   "Trulli"      "2"    "Wurz"        "9.15384615384615" "19"   
## 1998    "15"   "Trulli"      "0"    "Coulthard"   "7.04347826086957" "16"   
## 1999    "15"   "Panis"       "2"    "Alesi"       "7.79166666666667" "17"   
## 2000    "16"   "Zonta"       "2"    "Burti"       "9.73913043478261" "17"   
## 2001    "18"   "Button"      "1"    "Montoya"     "7.38461538461539" "15"   
## 2002    "17"   "Button"      "1"    "Barrichello" "8.08695652173913" "16"   
## 2003    "19"   "Wilson"      "0"    "Gene"        "9.125"            "17"   
## 2004    "17"   "Heidfeld"    "0"    "Gene"        "9.8"              "18"   
## 2005    "19"   "Karthikeyan" "0"    "De La Rosa"  "7.08"             "14"   
## 2006    "17"   "Coulthard"   "2"    "Raikkonen"   "8.33333333333333" "17"   
## 2007    "18"   "Sato"        "2"    "Nakajima"    "10.1153846153846" "17"   
## 2008    "17"   "Piquet"      "4"    "Davidson"    "10.2272727272727" "19"   
## 2009    "14"   "Trulli"      "3"    "Badoer"      "8.32"             "18"   
## 2010    "21"   "Di Grassi"   "2"    "Klien"       "11.962962962963"  "22"   
## 2011    "23"   "D’Ambrosio"  "0"    "De la Rosa"  "13.4827586206897" "24"   
## 2012    "23"   "Kovalainen"  "8"    "Vettel"      "16.16"            "23"   
## 2013    "20"   "Chilton"     "6"    "Webber"      "14.3913043478261" "20"   
## 2014    "21"   "Kobayashi"   "5"    "Hamilton"    "13.9130434782609" "21"   
## 2015    "17"   "Sainz"       "3"    "Rosberg"     "12.5238095238095" "19"   
## 2016    "22"   "Ericsson"    "2"    "Vettel"      "13.76"            "22"   
## 2017    "18"   "Ericsson"    "0"    "Giovinazzi"  "10.1666666666667" "19"   
## 2018    "19"   "Leclerc"     "4"    "Hamilton"    "12.65"            "19"   
## 2019    "19"   "Kubica"      "4"    "Leclerc"     "12.95"            "19"   
## 2020    "20"   "Raikkonen"   "2"    "Hamilton"    "11.304347826087"  "20"   
## General "66"   "Trulli"      "0"    "Tarquini"    "21.695652173913"  "88"   
##         Max out Name   Min out Min out Name   Mean out          
## 1994    "Katayama"     "0"     "Suzuki"       "5.45"            
## 1995    "Salo"         "0"     "Lavaggi"      "7.11428571428571"
## 1996    "Panis"        "0"     "Fisichella"   "6.375"           
## 1997    "Irvine"       "2"     "Wurz"         "9.15384615384615"
## 1998    "Coulthard"    "1"     "Hakkinen"     "7.04347826086957"
## 1999    "Irvine"       "2"     "Frentzen"     "7.79166666666667"
## 2000    "Barrichello"  "1"     "Mazzacane"    "9.73913043478261"
## 2001    "R.Schumacher" "0"     "Zonta"        "7.38461538461539"
## 2002    "Villeneuve"   "0"     "Davidson"     "8.08695652173913"
## 2003    "Montoya"      "1"     "Baumgartner"  "9.125"           
## 2004    "Alonso"       "0"     "Glock"        "9.8"             
## 2005    "R.Schumacher" "0"     "De la Rosa"   "7.08"            
## 2006    "Fisichella"   "0"     "Montagny"     "8.33333333333333"
## 2007    "Fisichella"   "0"     "Speed"        "10.1153846153846"
## 2008    "Hamilton"     "1"     "Sato"         "10.2272727272727"
## 2009    "Alonso"       "0"     "Badoer"       "8.32"            
## 2010    "Alonso"       "0"     "Yamamoto"     "11.962962962963" 
## 2011    "Button"       "0"     "De La Rosa"   "13.4827586206897"
## 2012    "Alonso"       "1"     "Karthikeyan"  "16.16"           
## 2013    "Raikkonen"    "3"     "Chilton"      "14.3913043478261"
## 2014    "Hamilton"     "0"     "Stevens"      "13.9130434782609"
## 2015    "Ricciardo"    "1"     "Merhi"        "12.5238095238095"
## 2016    "Verstappen"   "0"     "Pérez"        "13.76"
## 2017    "Hamilton"     "0"     "Di Resta"     "10.1666666666667"
## 2018    "Ricciardo"    "6"     "Sirotkin"     "12.65"
## 2019    "Vettel"       "2"     "Russell"      "12.95"
## 2020    "Leclerc"      "1"     "Fittipaldi"   "11.304347826087"
## General "Barrichello"  "0"     "Ratzenberger" "21.695652173913"
```

```
print("Degree in:")
```

```
## [1] "Degree in:"
```

```
sort(degree(g, mode = "in"), decreasing = TRUE)
```

```
##        Trulli   Barrichello        Button         Massa    Verstappen
##            66            57            57            57            55
##    Fisichella     Raikkonen        Alonso    De La Rosa      Heidfeld
##            55            53            53            50            49
##     Coulthard        Webber     Ricciardo         Perez         Panis
##            46            44            44            43            42
##         Glock   Karthikeyan         Sutil      Grosjean     Magnussen
##            42            42            42            42            41
##    Hulkenberg        Irvine          Salo         Diniz    Schumacher
##            41            40            39            39            38
##    Villeneuve        Kubica       Rosberg        Vettel          Sato
##            38            38            37            36            35
##        Liuzzi     Kobayashi     Maldonado      Frentzen      Hamilton
##            35            35            35            34            34
##      Ericsson    Kovalainen        Bottas      Katayama          Wurz
##            34            33            33            31            31
##       Herbert         Senna         Kvyat        Badoer  R.Schumacher
##            30            29            29            28            28
##        Petrov           Pic         Sainz          Ocon      Yamamoto
##            28            28            28            26            25
##     Gutierrez       Chilton      Wehrlein         Alesi      Hakkinen
##            25            25            25            24            24
##         Zonta         Buemi      Di Resta       Bianchi        Palmer
##            24            24            24            24            24
##        Stroll   Alguersuari    D’Ambrosio        Vergne          Nasr
##            24            23            23            23            23
##         Gasly      Haryanto       Leclerc      Nakajima     Di Grassi
##            23            22            22            21            21
##        Nakano      Monteiro     Vandoorne       Russell          Lamy
##            20            20            20            20            19
##        Takagi        Wilson        Albers        Piquet       Stevens
##            19            19            19            19            19
##    Giovinazzi          Hill   Baumgartner Van Der Garde         Albon
##            19            18            18            18            18
##        Latifi       Martini       Brabham       Zanardi        Rosset
##            18            17            17            17            17
##     Mazzacane      Davidson         Bruni         Klien       Hartley
##            17            17            17            17            17
##      Sirotkin        Norris    Fittipaldi          Gene      Bourdais
##            17            17            16            16            16
##      Chandhok         Merhi        Berger      Blundell         Inoue
##            16            16            15            15            15
##    Winkelhock    Morbidelli         Tuero         Yoong         Speed
##            15            14            14            14            14
##    De Cesaris    Montermini       Montoya        Firman      Da Matta
##            13            13            13            13            13
##         Rossi       Brundle        Moreno       Marques       Marquez
##            13            12            12            12            12
##  M.Schumacher      Pizzonia         Comas       Bernard      Alboreto
##            11            11            10            10            10
##      Bernoldi       Pantano        Gounon         Burti        McNish
##            10            10             9             9             9
##    Friesacher       Beretta        Gachot      Boullion      Doornbos
##             9             8             8             8             8
##    D’ambrosio           Ide        Suzuki        Dalmas Schiattarella
##             8             7             6             6             6
##      Deletraz      Sarrazin         Lehto         Kiesa    Wendlinger
##             6             6             5             5             4
##         Papis       Lavaggi      Montagny        Alliot       Mansell
##             4             4             4             3             3
##        Larini       Fontana          Enge        Aitken         Adams
##             3             3             3             3             2
##       Lagorce         Pérez  Ratzenberger          Noda    De la Rosa
##             2             2             1             1             1
##      Tarquini
##             0
```

```r
print("Degree out:")
```

```
## [1] "Degree out:"
```

```r
sort(degree(g, mode = "out"), decreasing = TRUE)
```

```
##   Barrichello     Raikkonen        Alonso         Massa        Button
##            88            81            81            80            77
##    Verstappen     Coulthard    Schumacher      Hamilton        Vettel
##            67            66            63            63            63
##        Webber       Rosberg         Panis    Fisichella         Perez
##            62            62            61            61            61
##    Hulkenberg      Heidfeld     Magnussen        Trulli     Ricciardo
##            59            57            56            55            55
##  R.Schumacher      Grosjean    Villeneuve         Sutil          Salo
##            54            53            52            49            46
##      Frentzen     Maldonado        Bottas        Irvine          Sato
##            44            44            44            41            41
##    Kovalainen         Kvyat       Herbert    De La Rosa     Kobayashi
##            41            40            39            37            37
##      Di Resta         Alesi      Katayama        Kubica         Sainz
##            37            36            36            36            36
##      Hakkinen        Liuzzi         Diniz       Montoya      Ericsson
##            35            34            33            32            32
##          Hill  M.Schumacher         Glock        Vergne          Wurz
##            31            30            30            30            29
##         Buemi        Petrov          Ocon   Alguersuari         Senna
##            29            29            28            26            26
##       Leclerc     Gutierrez        Stroll        Berger      Blundell
##            26            25            25            24            24
##         Gasly      Pizzonia         Zonta          Nasr        Norris
##            24            23            22            21            21
##       Brundle         Albon    Morbidelli         Klien      Bernoldi
##            20            20            18            18            17
##     Vandoorne      Nakajima        Piquet        Badoer        Palmer
##            17            16            16            15            15
##    Giovinazzi       Zanardi       Russell      Da Matta      Monteiro
##            15            14            14            12            12
##         Comas       Bernard       Mansell        Wilson         Lehto
##            11            11            11            11            10
##          Gene      Doornbos         Speed           Pic      Wehrlein
##            10            10            10            10            10
##    Fittipaldi       Martini    Montermini      Boullion        Rosset
##             9             9             9             9             9
##        Nakano   Karthikeyan      Bourdais       Hartley          Lamy
##             9             9             9             9             8
##    Wendlinger        Takagi      Davidson        Suzuki        Larini
##             8             8             8             7             7
##        Firman         Bruni       Chilton       Marques         Burti
##             7             7             7             6             6
##        McNish   Baumgartner        Albers     Di Grassi       Bianchi
##             6             6             6             6             6
##      Sirotkin        Moreno         Inoue      Yamamoto    Winkelhock
##             6             5             5             5             5
##       Beretta        Gachot         Papis       Fontana    De la Rosa
##             4             4             4             4             4
##    D’Ambrosio Van Der Garde         Rossi       Brabham    De Cesaris
##             4             4             4             3             3
## Schiattarella     Mazzacane      Chandhok      Haryanto        Latifi
##             3             3             3             3             3
##      Alboreto      Deletraz      Tarquini         Tuero      Sarrazin
##             2             2             2             2             2
##          Enge       Pantano    Friesacher    D’ambrosio       Stevens
##             2             2             2             2             2
##        Aitken        Gounon        Alliot        Dalmas       Lavaggi
##             2             1             1             1             1
##       Marquez         Yoong         Kiesa           Ide         Merhi
##             1             1             1             1             1
##  Ratzenberger         Adams       Lagorce          Noda      Montagny
##             0             0             0             0             0
##         Pérez
##             0
```

Watching the table we can see that in some season the minimum degree is 0. Talking about the in degree we can see that in the most of the years the least overtook is a top driver (sometimes even with 0 overtakes taken). Probably this could mean that in that year he had a very fast car and he was able to dominate or had a car which was in the middle of other cars (talking about performance) but was not close enough to battle with them. Watching the near column of the most overtaken we can see that every year this column has a name of a mid-low field driver. I was expecting this because those are the zones of the grid in which happen the most of the battles during the races because both drivers and cars have similar performances. Moving to the out degree the one that overtook less in every year except one was a low field driver, which can easily be explained because there is always a team with a slow car (compared to others) and because of that has a lot of struggle in overtaking. The only exception is Hakkinen in the 1998 season, in which he has a very strong car that permitted him to stay out of battles. on the other hand, that year, we have Coulthard as the one who made more overtakes which is a sort of contraddiction because Coulthard and Hakkinen drove for the same team. The other seasons the driver with most overtakes is a mid field driver or a "follow-up" driver (a driver that fought for the title till the end of the season but had not the fastest car or a driver which drove for the second/third fastest team that year). Because in almost every year there is a fastest car and other teams try to catch up during the season or are few tenths behind, this makes it so the drivers of the best team have an easier life than others during the races. Instead the follow-up drivers need to fight for thei position overtaking more cars.

Analyzing how graphs evolved during the years we can see that in the last second half we have less 0 in the in degree minimum column. That means that the races are more fought. We can come to this conclusion also watching at the columns with the mean of degrees: during the years the mean of degrees has increased for both in and out degree. With this data we can say that the federation has done a great job in increasing the show that Formula 1 races bring.

Moving on the general graph it has obviously 0 in both in and out minimum degree. Watching the degree lists the first things that comes to my eyes is that the two lists have 5 drivers in common in the first 9. Thinking about it I can say that it makes a lot of sense because are drivers which are (or have been) often in the first half of the grid, but have never had a dominating season, this means they always had to fight for their positions. If we search the drivers that dominated the most of the season in analysis (Schumacher, Vettel, Hamilton) we can find them in the top of the degree out list (they overtook a lot), but to find them in the degree in list we need to go down to the 25 position to find the first of them, which is Schumacher.

Betweenness centrality

```r
# Betweenness
cen_bet <- matrix(NA, nrow = length(g_yearly), ncol = 5)
colnames(cen_bet) <- c("Max", "Max Name", "Min", "Min Name", "Mean")
rownames(cen_bet) <- names(g_yearly)
for(year in 1994:2020) {
  cen_bet[as.character(year), "Max"] <- max(betweenness(g_yearly[[as.character(year)]],
                                                        directed = TRUE))
  cen_bet[as.character(year), "Max Name"] <- names(sort(betweenness(g_yearly[[as.character(year)]],
                                                                    directed = TRUE), decreasing = TRUE
  cen_bet[as.character(year), "Min"] <- min(betweenness(g_yearly[[as.character(year)]],
                                                        directed = TRUE))
  cen_bet[as.character(year), "Min Name"] <- names(sort(betweenness(g_yearly[[as.character(year)]],
                                                                    directed = TRUE)))[[1]]
  cen_bet[as.character(year), "Mean"] <- mean(betweenness(g_yearly[[as.character(year)]],
                                                          directed = TRUE))
}
cen_bet <- rbind(cen_bet, c(max(betweenness(g, directed = TRUE)), names(sort(
                      betweenness(g, directed = TRUE), decreasing = TRUE))[[1]],
                            min(betweenness(g, directed = TRUE)), names(sort(
                              betweenness(g, directed = TRUE)))[[1]],
                            mean(betweenness(g, directed = TRUE))))
rownames(cen_bet) <- c(names(g_yearly), "General")
cen_bet
```

```
##         Max                Max Name     Min                  Min Name
## 1994    "206.895812031106" "Katayama"   "0"                  "Schumacher"
## 1995    "189.735076103807" "Blundell"   "0"                  "Mansell"
## 1996    "60.199025974026"  "Panis"      "0"                  "Fisichella"
## 1997    "74.291539016539"  "Irvine"     "0"                  "Fontana"
## 1998    "90.0147186147186" "Alesi"      "0"                  "Coulthard"
## 1999    "88.5035353535353" "Hakkinen"   "1.64285714285714"   "Sarrazin"
## 2000    "56.4088383838384" "Diniz"      "0"                  "Burti"
## 2001    "100.722691197691" "Button"     "0"                  "Zonta"
## 2002    "96.6569805194805" "Villeneuve" "0"                  "Davidson"
## 2003    "53.7537424993307" "Coulthard"  "0"                  "Gene"
## 2004    "45.4357503607504" "Massa"      "0"                  "Glock"
## 2005    "72.4365079365079" "Villeneuve" "0"                  "De la Rosa"
## 2006    "67.6869855119081" "Monteiro"   "0"                  "Montagny"
## 2007    "60.6794184573596" "Sato"       "0"                  "Speed"
## 2008    "67.7709679209679" "Nakajima"   "1.33333333333333"   "Sato"
## 2009    "42.9688311688312" "Alonso"     "0"                  "Badoer"
## 2010    "60.9642588405746" "Hamilton"   "0"                  "Yamamoto"
## 2011    "52.2203415435613" "Perez"      "0"                  "De La Rosa"
## 2012    "77.4685045948204" "Kovalainen" "0"                  "D’ambrosio"
## 2013    "32.0168195108985" "Maldonado"  "0.4875"             "Kovalainen"
## 2014    "33.3672599459364" "Vergne"     "0"                  "Stevens"
## 2015    "38.0170634920635" "Rossi"      "0.0714285714285714" "Merhi"
## 2016    "33.2766193301023" "Massa"      "0"                  "Pérez"
## 2017    "39.2550976800977" "Palmer"     "0"                  "Giovinazzi"
## 2018    "21.7801587301587" "Verstappen" "0.624242424242424"  "Hartley"
## 2019    "27.8638972138972" "Albon"      "0.424242424242424"  "Russell"
## 2020    "54.0602036852037" "Raikkonen"  "0"                  "Hulkenberg"
## General "1720.0825458694"  "Panis"      "0"                  "Ratzenberger"
##         Mean
## 1994    "37.825"
## 1995    "37.3428571428571"
## 1996    "15.4583333333333"
## 1997    "17.8076923076923"
## 1998    "24.9565217391304"
## 1999    "21.0416666666667"
## 2000    "15.7391304347826"
## 2001    "21.8846153846154"
## 2002    "20.8260869565217"
## 2003    "11.9166666666667"
## 2004    "13.92"
## 2005    "18.92"
## 2006    "22.5555555555556"
## 2007    "13.6538461538462"
## 2008    "12.7727272727273"
## 2009    "15.48"
## 2010    "14.5555555555556"
## 2011    "14.8965517241379"
## 2012    "11.6"
## 2013    "9.60869565217391"
## 2014    "9"
## 2015    "10.1428571428571"
## 2016    "11.28"
## 2017    "11.2083333333333"
## 2018    "8.05"
## 2019    "7.75"
## 2020    "12.8260869565217"
## General "269.44213467297"
```

```
sort(betweenness(g, directed = TRUE), decreasing = TRUE)
```

```
##         Panis      Hamilton     Raikkonen   Barrichello     Coulthard
##  1720.0825459  1584.6548422  1505.9884274  1307.3357781  1238.6218687
##    Fisichella      Heidfeld     Magnussen    Villeneuve         Diniz
##  1190.0592456  1168.6476686  1156.7985890  1051.0255536  1004.7534230
##      Katayama          Salo        Kubica        Alonso        Trulli
##   983.5046090   981.6979186   924.9876515   920.4487012   883.3626613
##    De La Rosa        Webber        Irvine         Massa    Verstappen
##   846.5126375   824.4385142   812.9591122   778.4328318   760.7686647
##       Herbert        Bottas          Sato        Badoer         Rossi
##   748.3314886   678.0374105   667.7130622   619.9736812   591.4706097
##        Liuzzi        Button      Hakkinen      Grosjean       Rosberg
##   583.5732674   569.5629154   555.6426188   548.9691532   542.4787229
##    Fittipaldi        Vettel    Schumacher      Frentzen      Monteiro
##   535.0742092   531.7318082   508.2921069   467.0255873   459.1568983
##         Glock       Zanardi   Karthikeyan         Comas       Martini
##   435.7751679   418.6726735   385.1493694   377.1782169   374.1525750
##          Lamy      Blundell      Bernoldi         Sutil        Rosset
##   318.8168983   318.5570074   314.7827291   313.4668594   300.2423253
##     Maldonado    Kovalainen          Wurz   Baumgartner         Zonta
##   287.0654780   266.8737261   250.2432090   235.2643712   223.5078346
##        Suzuki        Piquet       Brabham      Yamamoto          Gene
##   222.1288496   220.6699711   214.8456241   214.1865037   208.5154944
##         Buemi           Pic         Senna  R.Schumacher      Pizzonia
##   206.3205641   200.4941200   190.0080862   184.6858242   181.4190249
##    Winkelhock    Montermini         Bruni       Stevens         Klien
##   179.6753092   177.9805830   163.2739946   158.7095238   154.9549965
##       Bianchi          Nasr        Berger    D’ambrosio      Nakajima
##   154.9445165   154.9087033   152.4593603   151.3384732   149.0703229
##    Morbidelli        Norris         Kvyat    Hulkenberg      Davidson
##   147.9012017   133.8852822   133.4328717   131.5580344   125.0858755
##     Di Grassi       Bernard        Petrov        Palmer     Kobayashi
##   122.2420131   121.3664851   116.2339796   114.6555891   114.5499985
##   Alguersuari       Marques       Brundle      Da Matta     Gutierrez
##   110.8832141   109.5200609   104.6372505   103.3349584   101.4263757
##      Bourdais        Stroll          Hill       Montoya  M.Schumacher
##   100.1257248   100.1012136    99.6875881    98.1236594    96.9693123
##          Ocon        Takagi     Ricciardo         Alesi      Ericsson
##    95.6031373    94.8741602    92.8815931    91.7298272    87.6845209
##         Perez      Boullion      Di Resta      Doornbos    Giovinazzi
##    85.0610875    83.0765642    75.0714540    74.6165571    72.9081986
##      Chandhok         Speed        Wilson         Inoue      Haryanto
##    71.7440446    71.3106428    70.8345819    66.8756344    62.5258151
##         Lehto        Aitken Schiattarella        McNish        Vergne
##    62.2343872    57.6529889    56.7339454    54.4910492    52.6866200
##      Wehrlein       Hartley         Albon       Chilton        Nakano
##    52.4549470    52.4332196    51.9201388    49.8886115    48.2977848
##      Deletraz      Sirotkin       Leclerc        Gachot         Burti
##    48.2611713    43.8328389    41.9628415    41.7365682    41.1222756
##    Friesacher     Vandoorne        Firman       Mansell     Mazzacane
##    34.2635761    30.9087916    30.2619030    29.7526008    27.8800517
##        Latifi        Albers        Moreno       Beretta    De Cesaris
##    27.0947625    26.6746737    26.4124725    19.7442013    19.6866088
##      Sarrazin          Enge    D’Ambrosio        Gounon         Sainz
##    17.3505600    14.0037278    13.2461937    12.0514667    11.7055862
##    Wendlinger        Larini         Yoong           Ide       Pantano
##    10.7403150    10.0475505     9.9300955     8.8323222     7.1470946
##      Alboreto       Fontana    De la Rosa         Tuero         Papis
##     6.8601659     6.7268249     5.7700325     5.4042823     4.9599618
##         Kiesa       Lavaggi        Dalmas       Russell         Gasly
##     3.8410958     3.1666768     2.1356322     1.9928571     0.5083333
##        Alliot  Ratzenberger         Adams       Lagorce          Noda
##     0.4022989     0.0000000     0.0000000     0.0000000     0.0000000
##      Tarquini       Marquez      Montagny Van Der Garde         Merhi
##     0.0000000     0.0000000     0.0000000     0.0000000     0.0000000
##         Pérez
##     0.0000000
```

Starting with the minimum betweenness we can see that almost every year there is a 0 on the value table; looking at the names we can see that again almost every year there is a low field driver. There are a few exception of top drivers probably because they never had problems during the season, so they fought only against other top drivers or because they have dominated the season. The low field drivers have probably the same reason but reversed: they had the slowest car so they only fought between team mates or other low field drivers that never had an exploit in that season. Moving on the highest values we can find all mid field drivers without any exception. This is exactly what I was excepting because these driver can have a very good week-end and battle even for the podium or can have a really bad week-end and battle with the low field drivers, that said it is easy to understand why they are there.

Opposed to the degree centrality analyzing how the betweenness changed during the years, it shows up that both the mean and the max value have decreased. The max value has a trend but it is has more up and downs than the mean, which has a very clear trend. This could mean that, combined with data from the degree analysis, the races are more fun to watch because there are more overtakes, but there is less chance to see an exploit both positive and negative one.

Page Rank

```
# PageRank
cen_pr <- matrix(NA, nrow = length(g_yearly), ncol = 5)
colnames(cen_pr) <- c("Max", "Max Name", "Min", "Min Name", "Mean")
rownames(cen_pr) <- names(g_yearly)
for(year in 1994:2020) {
  cen_pr[as.character(year), "Max"] <- max(page_rank(g_yearly[[as.character(year)]])$vector)
  cen_pr[as.character(year), "Max Name"] <- names(sort(page_rank(
                                              g_yearly[[as.character(year)]])$vector,
                                              decreasing = TRUE))[[1]]
  cen_pr[as.character(year), "Min"] <- min(page_rank(g_yearly[[as.character(year)]])$vector)
  cen_pr[as.character(year), "Min Name"] <- names(sort(page_rank(
                                            g_yearly[[as.character(year)]])$vector))[[1]]
  cen_pr[as.character(year), "Mean"] <- mean(page_rank(g_yearly[[as.character(year)]])$vector)
}
cen_pr <- rbind(cen_pr, c(max(page_rank(g)$vector), names(sort(page_rank(g)$vector,
```

```r
                                                                 decreasing = TRUE))[[1]],
                           min(page_rank(g)$vector), names(sort(page_rank(g)$vector))[[1]],
                           mean(page_rank(g)$vector)))
rownames(cen_pr) <- c(names(g_yearly), "General")
cen_pr
```

```
##         Max                  Max Name      Min                    Min Name
## 1994    "0.139195008282398"  "Brabham"     "0.00588802840627531"  "Schumacher"
## 1995    "0.0794625227477327" "Diniz"       "0.00480325474999113"  "Tarquini"
## 1996    "0.23747056796003"   "Badoer"      "0.00962177820487493"  "Hill"
## 1997    "0.0839437128717569" "Katayama"    "0.0114599281060625"   "Wurz"
## 1998    "0.163239236241806"  "Tuero"       "0.00652173913043479"  "Coulthard"
## 1999    "0.101740073156753"  "Takagi"      "0.00828574014244531"  "Coulthard"
## 2000    "0.122052130893838"  "Diniz"       "0.0113046716432908"   "Hakkinen"
## 2001    "0.176063911079887"  "Button"      "0.0074411809771178"   "Montoya"
## 2002    "0.108845577557754"  "Webber"      "0.00729336127626713"  "Barrichello"
## 2003    "0.162472960420796"  "Baumgartner" "0.00625"              "Gene"
## 2004    "0.0882043742719386" "Bruni"       "0.00719867050322328"  "Gene"
## 2005    "0.125188355616344"  "Karthikeyan" "0.00632717466343217"  "De La Rosa"
## 2006    "0.113674519767524"  "Monteiro"    "0.00936736916043813"  "Doornbos"
## 2007    "0.071836795727414"  "Sato"        "0.0136171864377481"   "Nakajima"
## 2008    "0.106807603100699"  "Nakajima"    "0.0221426977681004"   "Massa"
## 2009    "0.0672736010208062" "Trulli"      "0.0207058736506989"   "Kobayashi"
## 2010    "0.149344129026286"  "Di Grassi"   "0.0112410992412698"   "Klien"
## 2011    "0.102814835001146"  "D’Ambrosio"  "0.00575269522109456"  "De la Rosa"
## 2012    "0.156640662906257"  "Pic"         "0.012514401205841"    "Vettel"
## 2013    "0.105195658426872"  "Pic"         "0.0130712366886711"   "Vettel"
## 2014    "0.116081287009036"  "Ericsson"    "0.0137576522725523"   "Hamilton"
## 2015    "0.159903958702613"  "Stevens"     "0.0106754191814214"   "Rosberg"
## 2016    "0.0979216078556183" "Wehrlein"    "0.00839812701972212"  "Hamilton"
## 2017    "0.0968337583161954" "Ericsson"    "0.0067130731538153"   "Giovinazzi"
## 2018    "0.0845692088798186" "Leclerc"     "0.0107579847185872"   "Raikkonen"
## 2019    "0.0966849904229519" "Kubica"      "0.0103647494771332"   "Bottas"
## 2020    "0.10642445418119"   "Magnussen"   "0.00984893114266697"  "Hamilton"
## General "0.024066015172028"  "Trulli"      "0.000991460406764485" "Tarquini"
##         Mean
## 1994    "0.025"
## 1995    "0.0285714285714286"
## 1996    "0.0416666666666667"
## 1997    "0.0384615384615385"
## 1998    "0.0434782608695652"
## 1999    "0.0416666666666667"
## 2000    "0.0434782608695652"
## 2001    "0.0384615384615385"
## 2002    "0.0434782608695652"
## 2003    "0.0416666666666667"
## 2004    "0.04"
## 2005    "0.04"
## 2006    "0.037037037037037"
## 2007    "0.0384615384615385"
## 2008    "0.0454545454545454"
## 2009    "0.04"
## 2010    "0.037037037037037"
## 2011    "0.0344827586206897"
## 2012    "0.04"
## 2013    "0.0434782608695652"
## 2014    "0.0434782608695652"
## 2015    "0.0476190476190476"
## 2016    "0.04"
## 2017    "0.0416666666666667"
## 2018    "0.05"
## 2019    "0.05"
## 2020    "0.0434782608695652"
## General "0.0062111801242236"
```

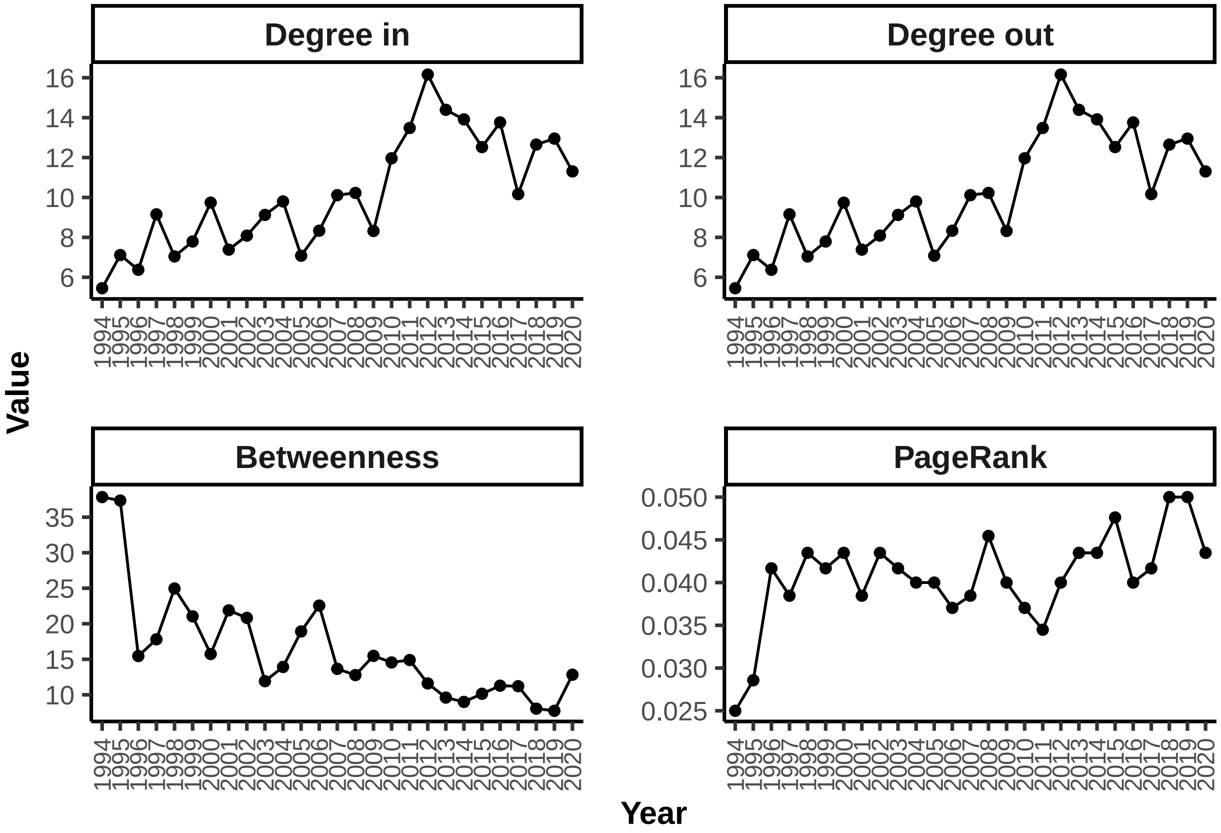
```
sort(page_rank(g)$vector, decreasing = TRUE)
```

```
##        Trulli        Button         Diniz      Ericsson     Magnussen
##  0.0240660152  0.0234865776  0.0213160408  0.0203622362  0.0163500348
##    Fisichella      Grosjean        Alonso   Barrichello         Glock
##  0.0159075015  0.0154559097  0.0147372715  0.0146186349  0.0143073114
##         Massa      Heidfeld       Stevens       Brabham         Perez
##  0.0139544735  0.0135850740  0.0132651014  0.0131258549  0.0126702275
##     Raikkonen    Kovalainen    De La Rosa          Sato         Sutil
##  0.0121260679  0.0119055922  0.0117111788  0.0116203541  0.0114793235
##     Ricciardo        Webber       Chilton         Merhi         Panis
##  0.0111279837  0.0110984196  0.0110500620  0.0108355237  0.0106956774
##        Liuzzi    Verstappen        Irvine     Maldonado     Coulthard
##  0.0105543916  0.0103709504  0.0103285914  0.0102659617  0.0102101447
##           Pic   Karthikeyan          Salo    Hulkenberg     Kobayashi
##  0.0101082648  0.0099690519  0.0098345070  0.0097694513  0.0097413060
##      Katayama        Bottas       Rosberg    Villeneuve Van Der Garde
##  0.0097350914  0.0094092651  0.0093689224  0.0092386669  0.0089251490
##      Monteiro        Badoer      Frentzen          Nasr        Kubica
##  0.0087827112  0.0085456128  0.0084334916  0.0081123856  0.0080165416
##        Moreno       Bianchi      Wehrlein        Albers     Gutierrez
##  0.0079453395  0.0079339090  0.0078025202  0.0075702855  0.0075568307
##         Kvyat        Palmer       Herbert          Wurz         Sainz
##  0.0075060131  0.0074528487  0.0074337372  0.0070959800  0.0070862537
##        Stroll        Petrov         Senna    Schumacher       Martini
##  0.0068891774  0.0068757565  0.0067090758  0.0063784864  0.0063767972
##          Ocon      Yamamoto      Hamilton          Lamy        Dalmas
##  0.0061752517  0.0061589850  0.0061351876  0.0058485917  0.0056833233
##      Hakkinen        Vettel  R.Schumacher    D’Ambrosio         Gasly
##  0.0055975341  0.0055852163  0.0055763543  0.0055508072  0.0055359425
##       Russell         Inoue         Alesi      Sirotkin         Rossi
##  0.0054750241  0.0054418068  0.0054309772  0.0054181946  0.0053148081
##         Zonta          Gene        Alliot    Montermini   Baumgartner
##  0.0052864495  0.0051468948  0.0051073700  0.0050939652  0.0050767394
##        Takagi      Di Resta     Mazzacane        Vergne      Blundell
##  0.0050595616  0.0050553395  0.0050286957  0.0049124083  0.0048369417
##   Alguersuari          Noda      Haryanto         Comas     Vandoorne
##  0.0047231207  0.0047104526  0.0047035524  0.0046823201  0.0045852765
##       Leclerc         Buemi        Rosset    Giovinazzi     Di Grassi
##  0.0045688848  0.0045154977  0.0044087487  0.0043309169  0.0043028110
##       Bernard    Fittipaldi      Alboreto         Tuero         Bruni
##  0.0042991400  0.0041513831  0.0041245691  0.0039286356  0.0038779463
##      Chandhok       Marquez        Nakano       Hartley      Nakajima
##  0.0038380621  0.0036914746  0.0036590185  0.0035403472  0.0034270087
##        Wilson          Hill       Zanardi      Deletraz    Friesacher
##  0.0034085499  0.0034017849  0.0033760523  0.0033685252  0.0032902274
##         Speed    De Cesaris           Ide        Piquet        Gachot
##  0.0032549058  0.0031690345  0.0031646939  0.0030129681  0.0028805744
##         Klien    Morbidelli         Yoong      Bourdais        Norris
##  0.0028481923  0.0028392317  0.0028319218  0.0028001030  0.0026017619
## Schiattarella         Albon        Berger        Latifi       Lavaggi
##  0.0025939561  0.0025887061  0.0025402370  0.0025273691  0.0024278509
##        Suzuki      Davidson       Pantano        Gounon        Firman
##  0.0024231374  0.0024089390  0.0023164527  0.0022872026  0.0022828088
##       Brundle      Boullion      Bernoldi       Montoya         Papis
##  0.0022693681  0.0022573605  0.0022217735  0.0021995244  0.0020630184
##      Da Matta    Winkelhock       Beretta      Sarrazin  M.Schumacher
##  0.0020532397  0.0020268825  0.0020114099  0.0019845603  0.0019649889
##         Kiesa       Marques      Pizzonia         Burti         Lehto
##  0.0019582224  0.0019564283  0.0018906935  0.0017881398  0.0017617336
##        McNish      Doornbos       Lagorce  Ratzenberger      Montagny
##  0.0016800423  0.0015041129  0.0014841124  0.0014596924  0.0014453612
##    Wendlinger       Fontana    D’ambrosio          Enge         Adams
##  0.0012695726  0.0012156506  0.0012073939  0.0011976082  0.0011858009
##        Larini       Mansell    De la Rosa        Aitken         Pérez
##  0.0011615183  0.0011584699  0.0010988224  0.0010561396  0.0010382575
##      Tarquini
##  0.0009914604
```

As we can see the majority of drivers with the highest PageRank value during the years are mid/low field drivers because, like we have seen in the previous centrality analysis, these are the areas where most of overtakes happen. This means that those drivers has a high number of in-links and those link come from other nodes with high PageRank. On the other side, the drivers with the lowest values are drivers from the very top or from the very bottom of the grid because in those areas there are less overtakes which means less in-links. Analyzing the values of the general graph, instead, we can see that there are some differences because in this graph the in-links are influenced not only by the overtakes in one single season but in multiple seasons. This means that a driver that raced for 15 years probably has a higher PageRank than a driver that raced in only one season but its has been overtaken a lot. In fact we can see that the drivers with lower PageRank are drivers that raced for just few years mixed with drivers in the very top or very bottom of the grid. Instead between the ones with high values we can find both drivers from mid field and drivers that raced for most of the years of this sample.

Temporal analysis

```r
# Creating the plot to show the temporal trend of centralities
temporal_matrix <- matrix(NA, nrow = length(g_yearly), ncol = 4)
colnames(temporal_matrix) <- c("Degree in", "Degree out", "Betweenness", "PageRank")
rownames(temporal_matrix) <- names(g_yearly)

for(year in 1994:2020) {
  temporal_matrix[as.character(year), "Degree in"] <- cen_degree[as.character(year), "Mean in"]
  temporal_matrix[as.character(year), "Degree out"] <- cen_degree[as.character(year), "Mean out"]
  temporal_matrix[as.character(year), "Betweenness"] <- cen_bet[as.character(year), "Mean"]
  temporal_matrix[as.character(year), "PageRank"] <- cen_pr[as.character(year), "Mean"]
}

df <- as.data.frame(temporal_matrix)
df <- rownames_to_column(df, var = "Year")
df_long <- pivot_longer(
  df,
  cols = c("Degree in", "Degree out", Betweenness, PageRank),
  names_to = "Index",
  values_to = "Value"
)
df_long$Value <- as.numeric(df_long$Value)
df_long$Index <- factor(df_long$Index, levels = c("Degree in", "Degree out",
                                                  "Betweenness", "PageRank"))

ggplot(df_long, aes(x = Year, y = Value, group = Index)) +
  geom_line() +
  geom_point() +
  facet_wrap(~Index, ncol = 2, scales = "free_y", strip.position = "top", axes = "all") +
  scale_y_continuous(breaks = scales::pretty_breaks(n = 5)) +
  theme_classic(base_size = 14) +
  theme(strip.placement = "outside",
        panel.spacing = unit(1.5, "lines"),
        strip.text = element_text(size = 12, face = "bold"),
        axis.text.x = element_text(size = 9.3, angle = 90, vjust = 0.5, hjust = 1),
        axis.text.y = element_text(size = 10),
        axis.title.x = element_text(size = 12, face = "bold"),
        axis.title.y = element_text(size = 12, face = "bold"))
```



As we saw in the tables with numeric values, the graphs expose in a better way the change during the years. We can see a clear increasing trend in the Degree centrality until 2012 and then a little decrease. Talking about the PageRank we can still see an increasing trend, but it less clean, with more up and downs. On the other hand, we have a clear decreasing trend for the Betweenness centrality, with some outliers, but mostly in the early years of the sample.

This means that the action during the races has increased (there are more total overtakes), and it is more various. The Betweenness decreasing means that there are more paths that link two nodes which do not pass through a single node. So the drivers tend to overtake not only some drivers that have a similar potential, but sometimes overtake drivers from a different zone of the grid. That said, we can say that the work the FIA is doing to increase the show during the races is working and the teams reached a very high level of preparation allowing the drivers to make much more exploit.

Community analysis

```r
methods <- list(
  "Edge Betweenness" = cluster_edge_betweenness,
  "Fast Greedy" = cluster_fast_greedy,
  "Label Propagation" = cluster_label_prop,
  "Leading Eigenvector" = cluster_leading_eigen,
  "Louvain" = cluster_louvain,
  "Walktrap" = cluster_walktrap,
  "Spinglass" = function(graph) cluster_spinglass(graph, spins = 10),
  "Infomap" = cluster_infomap,
  "Optimal" = cluster_optimal
)

# Compute modularity for each method
results <- data.frame(Method = character(), Modularity = numeric(), stringsAsFactors = FALSE)

for (method in names(methods)) {
  tryCatch({
    # Detect communities
    communities <- methods[[method]](g_yearly[["1994"]])

    # Compute modularity
    modularity_value <- modularity(communities)

    # Store the result
    results <- rbind(results, data.frame(Method = method, Modularity = modularity_value))
  }, error = function(e) {
    # Handle any errors (e.g., if a method is not applicable)
    cat("Error with method:", method, "\n")
  })
}
```

```
## Error with method: Fast Greedy
```

```
## Error with method: Louvain
```

```r
# Sort results by modularity in decreasing order
results <- results[order(-results$Modularity), ]

# Print the results
results
```

```
##                 Method    Modularity
## 7              Optimal  2.272326e-01
## 5            Spinglass  2.012562e-01
## 3 Leading Eigenvector  1.984261e-01
## 4             Walktrap  9.403670e-02
## 6              Infomap  4.019022e-02
## 1    Edge Betweenness  8.921808e-03
## 2   Label Propagation -1.110223e-16
```

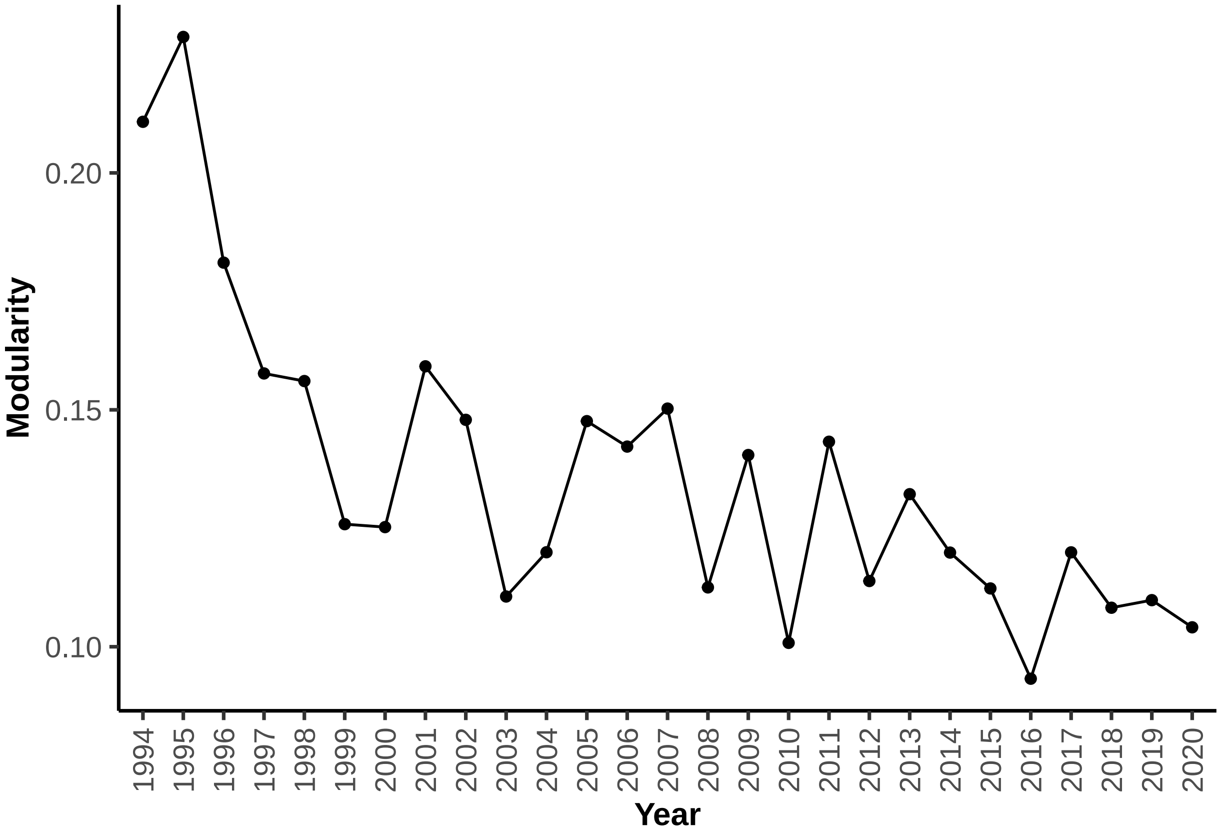
To analyze the communities, I decided to use the Spinglass algorithm because in the test that I have done above there was not a big difference between the Optimal and the Spinglass algorithms. So I decided to use the one who uses less time and resources to compute.

```r
# Community detection with Spinglass algorithm
modularity_general <- modularity(methods[["Spinglass"]](g))
modularity_yearly <- c()
for (year in 1994:2020) {
  modularity_yearly <- c(modularity_yearly, modularity(methods[["Spinglass"]](
                                             g_yearly[[as.character(year)]])))
}
names(modularity_yearly) <- names(g_yearly)
```

```
modularity_general
```

```
## [1] 0.1018801
```

```
# Plot the annual trend of modularity
modularity_df <- as.data.frame(modularity_yearly)
modularity_df <- rownames_to_column(modularity_df, var = "Year")
colnames(modularity_df)[2] <- "Modularity"
modularity_df$Modularity <- as.numeric(modularity_df$Modularity)

ggplot(modularity_df, aes(x = Year, y = Modularity, group = 1)) +
  geom_point() +
  geom_line() +
  scale_y_continuous(breaks = scales::pretty_breaks(n = 5)) +
  theme_classic(base_size = 14) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1),
        axis.title.x = element_text(size = 12, face = "bold"),
        axis.title.y = element_text(size = 12, face = "bold"))
```



As we can see from the data, the modularity value for every graph is pretty low. It is positive tho, so the number of edges between nodes of the same community is higher than the same value in a random graph. This means that drivers tend to form groups in which the overtakes happen more frequently inside than outside of the groups. But the low values means that this groups are not that defined (there are groups, but there are also a lot of edges that go from one group to another). Looking at the graph, this trend has decreased during the years, which is something that I was expecting. Because, as I said earlier talking about centrality, with the degree and PageRank centrality increasing and the betweenness decreasing, the number of paths that link two nodes has increased, so it is less common for a node to be in the majority of those paths.

Connectivity

```
# Finding giant component for each graph
giant_general <- c(max(components(g, mode="strong")$csize), 100*max(components(
                                                  g, mode="strong")$csize)/vcount(g))
names(giant_general) <- c("C Size", "C Size%")

giant_yearly <- matrix(NA, nrow = length(g_yearly), ncol = 2)
rownames(giant_yearly) <- names(g_yearly)
colnames(giant_yearly) <- c("C Size", "C Size%")
for (year in 1994:2020) {
  giant_yearly[as.character(year), "C Size"] <- max(components(g_yearly[[as.character(year)]],
                                                               mode="strong")$csize)
  giant_yearly[as.character(year), "C Size%"] <- 100*(giant_yearly[as.character(year),
                                                "C Size"]/vcount(g_yearly[[as.character(year)]]))
}

giant_general <- round(giant_general, 2)
giant_yearly <- round(giant_yearly, 2)

giant_general
```

```
##  C Size C Size%
##  151.00   93.79
```

```
giant_yearly
```

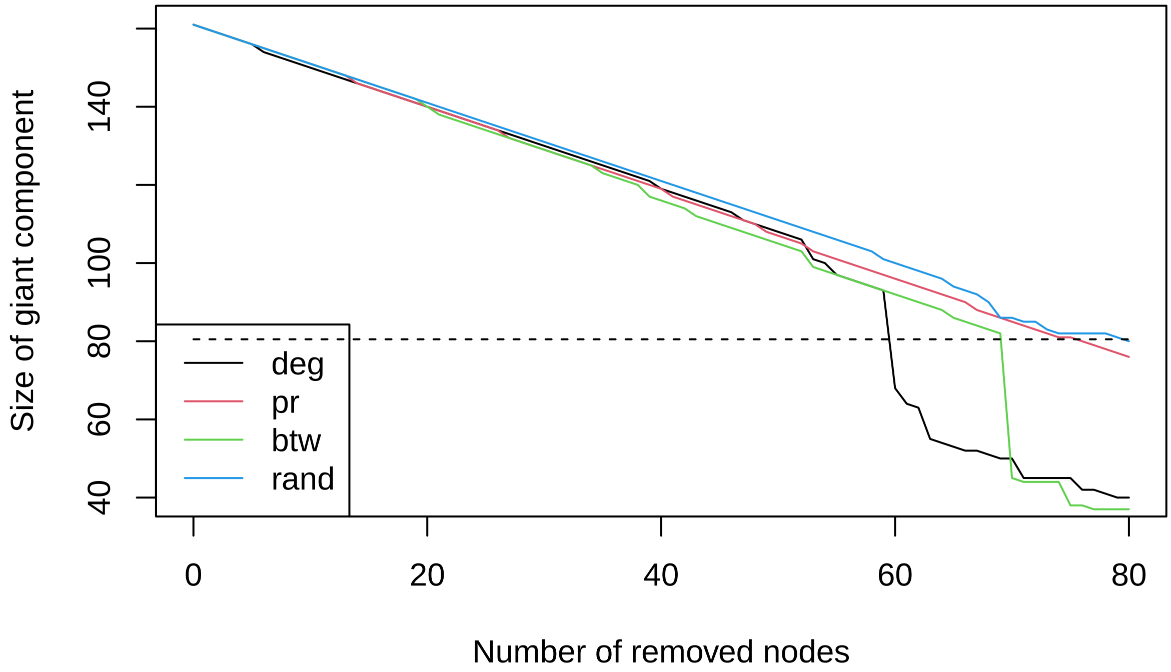
```
##      C Size C Size%
## 1994     29   72.50
## 1995     33   94.29
## 1996     19   79.17
## 1997     26  100.00
## 1998     22   95.65
## 1999     24  100.00
## 2000     23  100.00
## 2001     24   92.31
## 2002     22   95.65
## 2003     20   83.33
## 2004     23   92.00
## 2005     22   88.00
## 2006     26   96.30
## 2007     24   92.31
## 2008     22  100.00
## 2009     23   92.00
## 2010     25   92.59
## 2011     26   89.66
## 2012     25  100.00
## 2013     23  100.00
## 2014     22   95.65
## 2015     21  100.00
## 2016     24   96.00
## 2017     20   83.33
## 2018     20  100.00
## 2019     20  100.00
## 2020     23  100.00
```

We can see that the percentage of nodes in the giant strongly connected component is very high for every graph. Every year has at least 70% of nodes in the strongly connected component. This is a good news for the FIA because it means that the chain of overtakes reaches the majority of drivers. The result meets my expectations, especially after the result of modularity analysis. Because the communities exist but are not isolated, this means that there are a few extra-community edges which contribute to enlarge the connected component.

Resiliencce

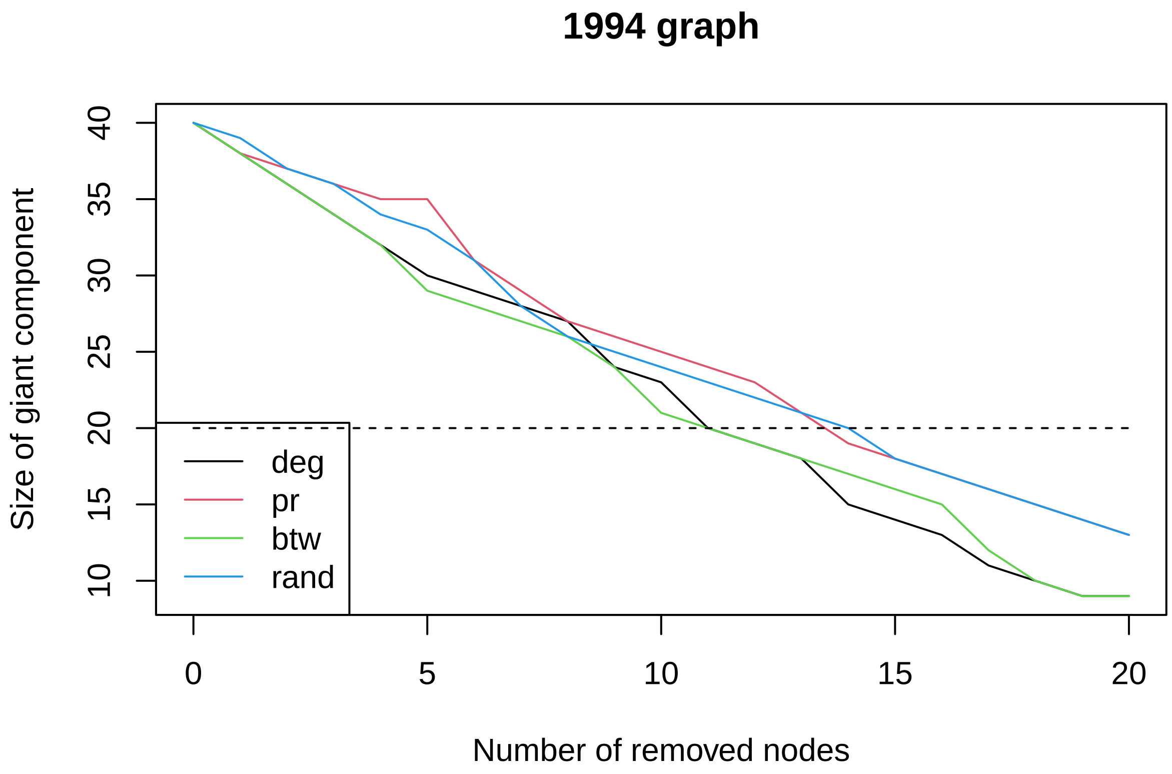
```
# percolation removes nodes from a graph and computes
# the size of the giant connected component
# INPUT
# g: graph to percolate
# size: number of nodes to remove
# d: removal vector
# OUTPUT
# giant: a vector with sizes of giant components when nodes are removed
percolate = function(g, size, d) {

  giant = vector()

  # initial size of giant component
  c = components(g)
  giant[1] = max(c$csize)

  # find vital nodes
  names(d) = 1:length(d)
  d = sort(d, decreasing=TRUE)
  vital = as.integer(names(d[1:size]))

  # compoute size of giant component after incremental removal
  for (i in 1:size) {
    c = components(delete_vertices(g, vital[1:i]))
    giant[i+1] = max(c$csize)
  }

  return(giant)

}

# percolation of general graph
size = vcount(g)/2
# random
```

```
rand = percolate(g, size, d = sample(V(g), size))
# degree
deg = percolate(g, size, d = degree(g))
# pagerank
pr = percolate(g, size, d=page_rank(g)$vector)
# betweenness
bet = percolate(g, size, d = betweenness(g))

plot(0:size, deg, type = "l", col=1,
     xlab="Number of removed nodes",
     ylab="Size of giant component")
lines(0:size, pr, col=2)
lines(0:size, bet, col=3)
lines(0:size, rand, col=4)
lines(0:size, rep(vcount(g)/2, size+1), lty=2)
legend(x = "bottomleft",
       legend = c("deg", "pr", "btw", "rand"), lty = 1, col = 1:4)
```

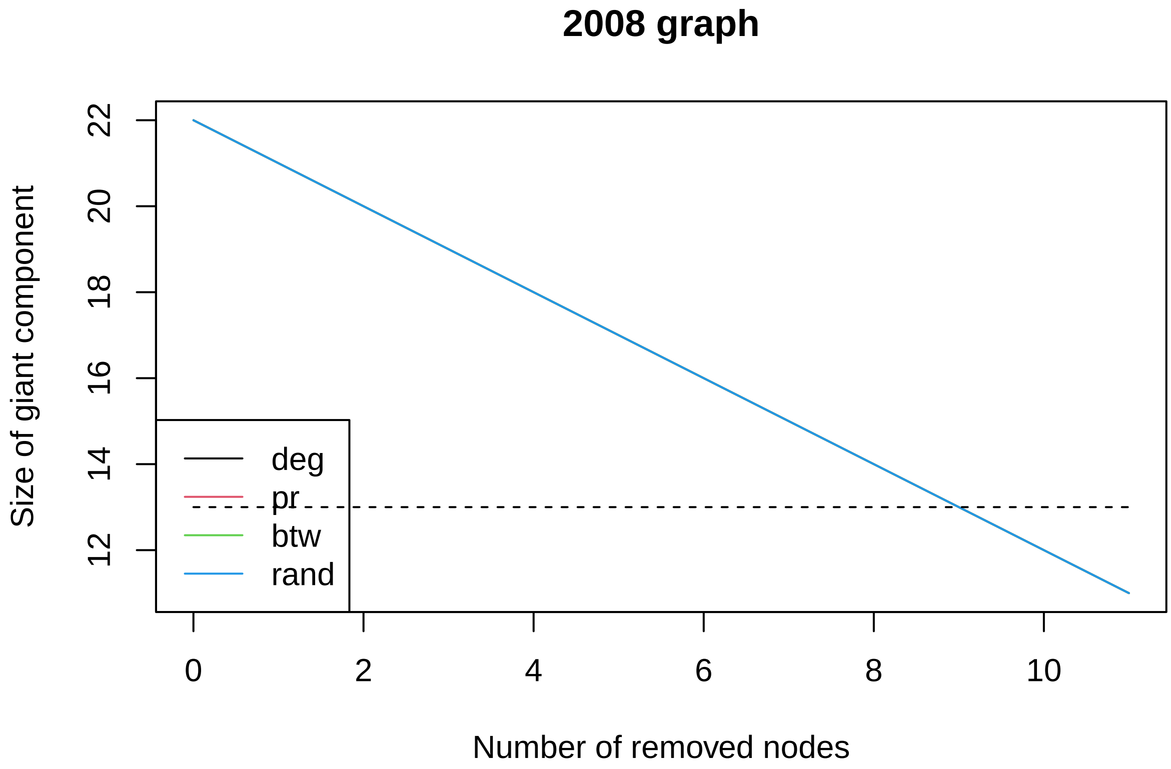


As we can see from the plot, the general graph is very resilient. In fact it would be heavily damaged only by an attack that follow the degree centrality and the betweenness centrality. Both of them start to be effective only after more than 50 nodes removed. In particular the degree attack is the first one to cause the damage, as we can see looking at the black line, after the removal of the 60th node causing a drop in the dimension of the giant component (from near 100 to near 70). The betweenness attack, instead, is effective only after the 70th node removed, but it causes a bigger drop in the dimension of the giant component (from 80 to near 40). To see if the annual graphs have the same resilience, I decided to analyze a sample of them. I did not analyze them all because we saw in the previous analysis that they are pretty similar. To choose the sample I took the graph with the least percentage of nodes in the giant component, one from the ones who have 100% of nodes in the giant component and one in the middle (with the percentage around 85%)

```
# percolation of one annual graph (1994) (72,5%)
size = vcount(g_yearly[["1994"]])/2
# random
rand = percolate(g_yearly[["1994"]], size, d = sample(V(g_yearly[["1994"]]), size))
# degree
deg = percolate(g_yearly[["1994"]], size, d = degree(g_yearly[["1994"]]))
# pagerank
pr = percolate(g_yearly[["1994"]], size, d=page_rank(g_yearly[["1994"]])$vector)
# betweenness
bet = percolate(g_yearly[["1994"]], size, d = betweenness(g_yearly[["1994"]]))

plot(0:size, deg, type = "l", col=1,
    xlab="Number of removed nodes",
    ylab="Size of giant component",
    main = "1994 graph")
lines(0:size, pr, col=2)
lines(0:size, bet, col=3)
lines(0:size, rand, col=4)
lines(0:size, rep(vcount(g_yearly[["1994"]])/2, size+1), lty=2)
legend(x = "bottomleft",
       legend = c("deg", "pr", "btw", "rand"), lty = 1, col = 1:4)
```

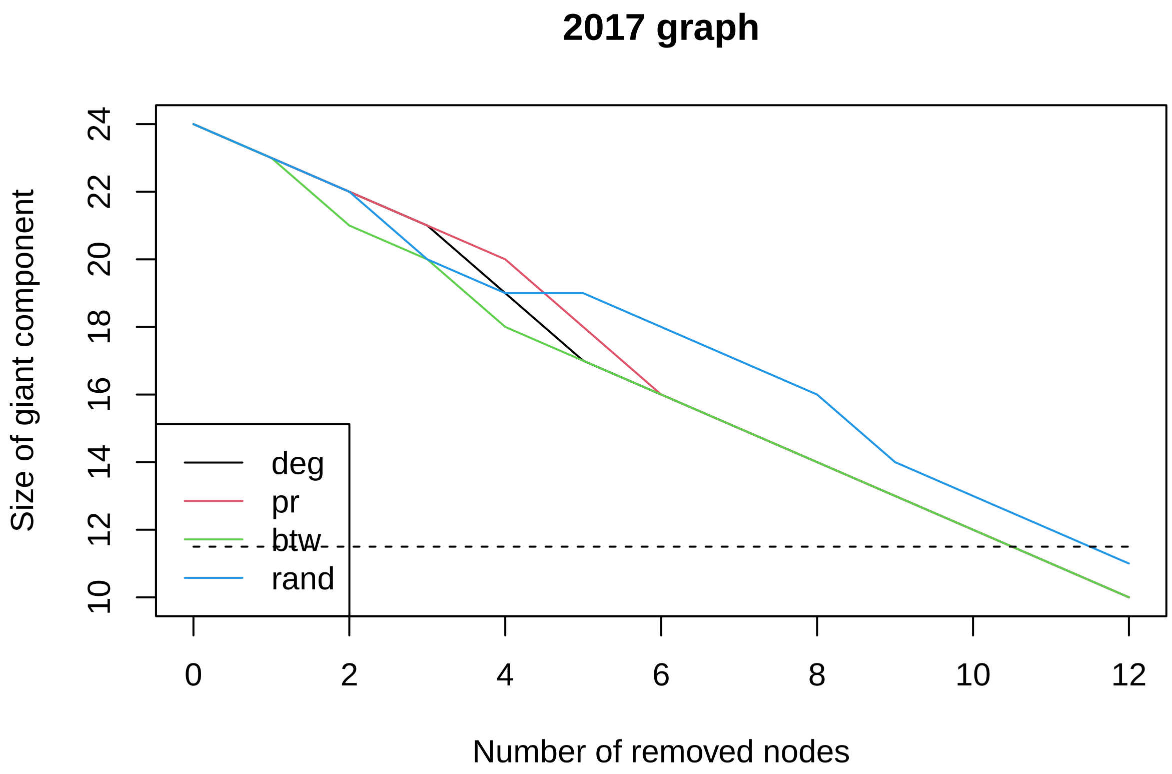


```r
# percolation of one annual graph (2008) (100%)
size = vcount(g_yearly[["2008"]])/2
# random
rand = percolate(g_yearly[["2008"]], size, d = sample(V(g_yearly[["2008"]]), size))
# degree
deg = percolate(g_yearly[["2008"]], size, d = degree(g_yearly[["2008"]]))
# pagerank
pr = percolate(g_yearly[["2008"]], size, d=page_rank(g_yearly[["2008"]])$vector)
# betweenness
bet = percolate(g_yearly[["2008"]], size, d = betweenness(g_yearly[["2008"]]))

plot(0:size, deg, type = "l", col=1,
    xlab="Number of removed nodes",
    ylab="Size of giant component",
    main = "2008 graph")
lines(0:size, pr, col=2)
lines(0:size, bet, col=3)
lines(0:size, rand, col=4)
lines(0:size, rep(vcount(g_yearly[["2007"]])/2, size+1), lty=2)
legend(x = "bottomleft",
       legend = c("deg", "pr", "btw", "rand"), lty = 1, col = 1:4)
```



```r
# percolation of one annual graph (2017) (83,33%)
size = vcount(g_yearly[["2017"]])/2
# random
```

```
rand = percolate(g_yearly[["2017"]], size, d = sample(V(g_yearly[["2017"]]), size))
# degree
deg = percolate(g_yearly[["2017"]], size, d = degree(g_yearly[["2017"]]))
# pagerank
pr = percolate(g_yearly[["2017"]], size, d=page_rank(g_yearly[["2017"]])$vector)
# betweenness
bet = percolate(g_yearly[["2017"]], size, d = betweenness(g_yearly[["2017"]]))

plot(0:size, deg, type = "l", col=1,
     xlab="Number of removed nodes",
     ylab="Size of giant component",
    main = "2017 graph")
lines(0:size, pr, col=2)
lines(0:size, bet, col=3)
lines(0:size, rand, col=4)
lines(0:size, rep(vcount(g_yearly[["2020"]])/2, size+1), lty=2)
legend(x = "bottomleft",
       legend = c("deg", "pr", "btw", "rand"), lty = 1, col = 1:4)
```



Looking at the plots we can see that these graph are even more resilient than the general. Analyzing them in order, we have the 1994 graph that is resilient, but suffers a little bit the attacks made following the degree and the betweenness centrality. There are no drops, but the lines of this two types of attacks are more sharp than the other two at the start. The PageRank attack starts with a lot of difficulty, even more than the random one at a certain point. Then has a little drop, after 5 nodes, reaching the effectiveness of the degree and the betweenness, but then it gets worse and reaches again the random attack. In the 2008 graph, as we could have expected, all the attacks has the same effectiveness because with a giant component that covers the totality of the graph every node means that there is only 1 node less than the previous removal. Last but not least, we have the 2017 graph, which is the most resilient, in fact the giant component (after the removal of half of the nodes) is just smaller than the half of the graph. It is interesting to see that after the first 2 removals the more effective attack is the random one, shortly reached by the betweenness one. With the PageRank that is the last to reach that level of effectiveness. In the second half of removals the organised attacks (Degree, PageRank and Betweenness) return to be more effective than the random.

Distances and Degree distribution

Mean distances and diameters

```
# Mean distance and diameter
distance_general <- c(mean_distance(g), diameter(g))
names(distance_general) <- c("Mean distance", "Diameter")

distance_yearly <- matrix(NA, nrow = length(g_yearly), ncol = 2)
rownames(distance_yearly) <- names(g_yearly)
colnames(distance_yearly) <- c("Mean distance", "Diameter")
for (year in 1994:2020) {
  distance_yearly[as.character(year), "Mean distance"] <- mean_distance(
                                                        g_yearly[[as.character(year)]])
  distance_yearly[as.character(year), "Diameter"] <- diameter(g_yearly[[as.character(year)]])
}
distance_general <- round(distance_general, 2)
distance_yearly <- round(distance_yearly, 2)

distance_general
```

```
## Mean distance      Diameter
##             3            11
```

```
distance_yearly
```

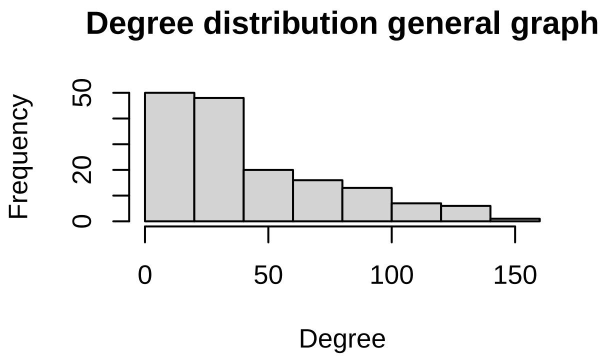
```
##      Mean distance Diameter
## 1994          2.32        5
## 1995          2.16        5
## 1996          1.83        4
## 1997          1.71        3
## 1998          2.19        5
## 1999          1.91        5
## 2000          1.72        5
## 2001          1.95        5
## 2002          1.99        6
## 2003          1.61        4
## 2004          1.63        4
## 2005          1.89        5
## 2006          1.90        5
## 2007          1.59        3
## 2008          1.61        4
## 2009          1.70        3
## 2010          1.60        4
## 2011          1.59        4
## 2012          1.48        5
## 2013          1.44        4
## 2014          1.43        4
## 2015          1.51        5
## 2016          1.49        4
## 2017          1.58        4
## 2018          1.42        4
## 2019          1.41        4
## 2020          1.58        4
```

Looking at the table, we can see that the mean distances are very low in every graph. In the general graph there is a mean distance of 3, which is very low considering that the graph has 161 nodes and almost 4000 edges. The diameter is a little bit longer, but not comparable to the total number of nodes. The annual graphs have a very low mean distance, in almost every year, but 2, the mean distance is less than 2. The diameters are between 3 and 6, with a majority of 4 and 5. After this analysis we can say that these graph are small worlds, because both the mean distances and the diameters are very low.

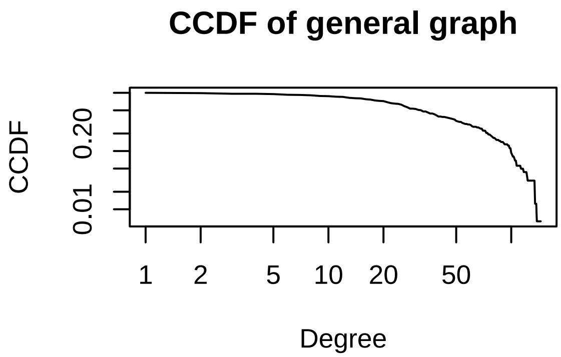
After having defined the small worlds, we can analyze the degree distribution to see if it is also a scale free network. To be so the degree distribution has to be a power law.

Degree distribution

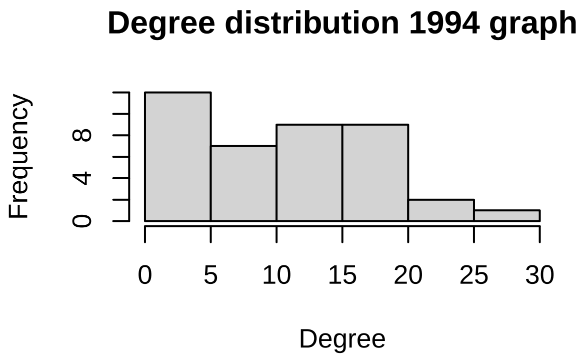
```r
# Cumulative distribution
ccdf = function(d) {
  n = length(d)
  max = max(d)
  p = rep(0, max)
  for (i in 1:length(p)) {
    p[i] = length(d[d >= i]) / n
  }
  return(p)
}

par(mfrow = c(2, 2))
# Degree distribution (general)
hist(degree(g), main = "Degree distribution general graph", xlab = "Degree")
# CCDF (general)
plot(ccdf(degree(g)), log = "xy", type = "l", main = "CCDF of general graph",
     xlab = "Degree", ylab = "CCDF")

# Degree distribution (1994)
hist(degree(g_yearly[["1994"]]), main = "Degree distribution 1994 graph",
     xlab = "Degree")
# CCDF (general)
plot(ccdf(degree(g_yearly[["1994"]])), log = "xy", type = "l", main = "CCDF of 1994 graph",
     xlab = "Degree", ylab = "CCDF")
```



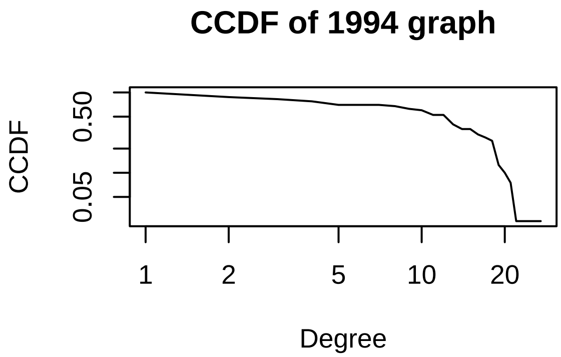
Degree distribution general graph



CCDF of general graph

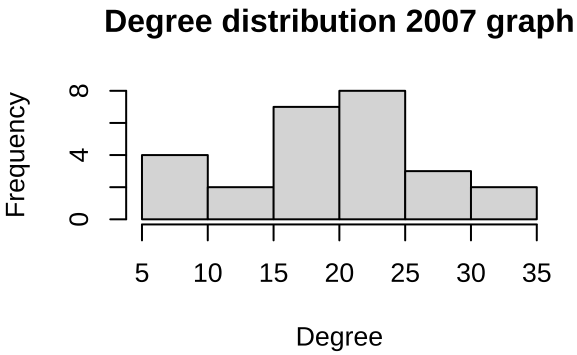


Degree distribution 1994 graph

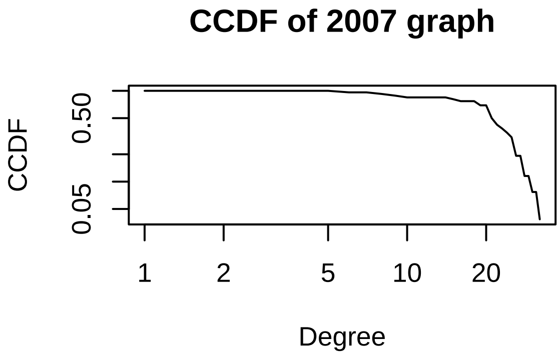


CCDF of 1994 graph

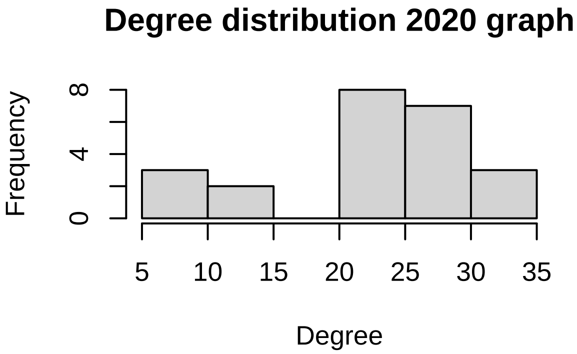
```
# Degree distribution (2007)
hist(degree(g_yearly[["2007"]]), main = "Degree distribution 2007 graph", xlab = "Degree")
# CCDF (general)
plot(ccdf(degree((g_yearly[["2007"]]))), log = "xy", type = "l", main = "CCDF of 2007 graph",
     xlab = "Degree", ylab = "CCDF")

# Degree distribution (2020)
hist(degree(g_yearly[["2020"]]), main = "Degree distribution 2020 graph", xlab = "Degree")
# CCDF (general)
plot(ccdf(degree(g_yearly[["2020"]])), log = "xy", type = "l", main = "CCDF of 2020 graph",
     xlab = "Degree", ylab = "CCDF")
```



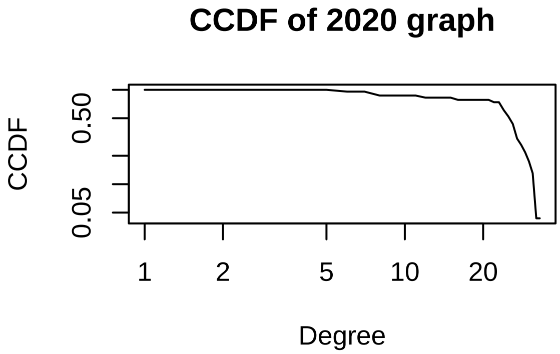
Degree distribution 2007 graph



CCDF of 2007 graph



Degree distribution 2020 graph



CCDF of 2020 graph

Looking at the plots we can say that the general graph has almost a degree distribution power law. It is not perfect, but it is something like that. On the other hand the annual graphs do not have a power law distribution. The graph of 1994 has a distribution that is right asymmetric, but not like a power law. While the other 2 analyzed graphs have a distribution which is slightly asymmetric, but definitely not a power law. These last two are more similar to a poisson distribution.

Conclusion

In conclusion we can say that the general graph and the annual graphs have some similarities, but also have some differences. Starting with the last things that was analyzed, we saw that the general graph is a small world network with a power law (almost) degree distribution, while the annual graphs are small world networks but definitely do not have power law degree distribution (some graphs are closer to a power law, others to a poisson). We saw that all graphs are pretty resilient, both the general and the annuals are more susceptible to degree and betweenness centrality, but the annual graphs tend to be slightly more resilient. The pagerank attack, instead, is the best handled by every graph, but still a little more effective than the random attack. Looking at the evolution over the years, we can see that the degree centrality and the pagerank increased, while the betweenness decreases. This means that the work the FIA is doing to increase the overtakes and the show of the Formula 1 is working. This is shown also by the modularity and communities: during the years the communities have become less isolated, there is still a division in groups of drivers that overtake each other more but it is less defined. In the last years the drivers tend to overtake more and drivers from all the grid.