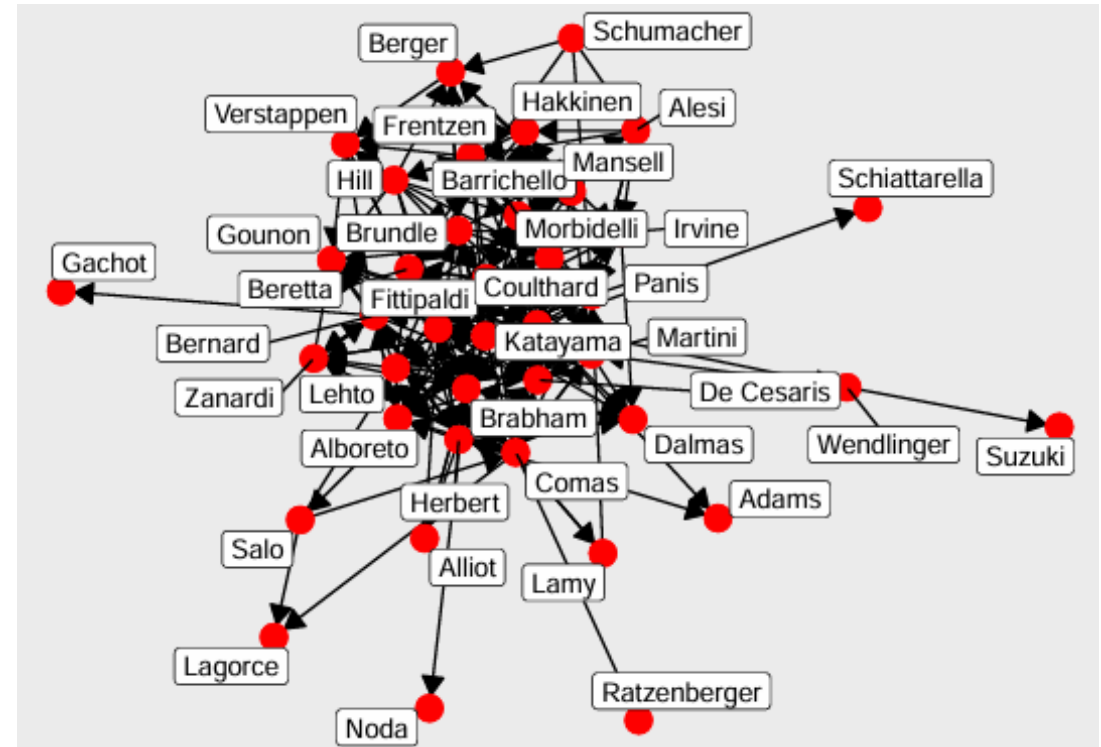# F1 OVERTAKES ANALYSIS

Advanced Data Science 2024,

Relato Alessandro

THE GRAPHS

The graphs represents the overtakes that happen during a season of Formula 1.

There are 26 graphs of the seasons from 1994 to 2020 and one general graph with all the seasons together.

All 27 are directed graph and every edge go from the overtaker to the overtaken.

Graph of the 1994 season

| Graph	# of nodes	# of edges	Density	Reciprocity
Mean	25,11	250,48	0,4378	0,4714
General	161	3493	0,1356	0,4689

CENTRALITY

I analyzed 3 type of centrality: degree, betweenness and PageRank.

The degree centrality analysis pointed out that the driver which was overtaken the most and the one who overtook the least were a mid-low field driver.

The driver that overtook the most is almost every year a mid field driver.

While the driver which was overtaken the least is or a top driver or a low-field driver.

During the years the degree centrality has increased till 2012 and then has decreased a little bit.
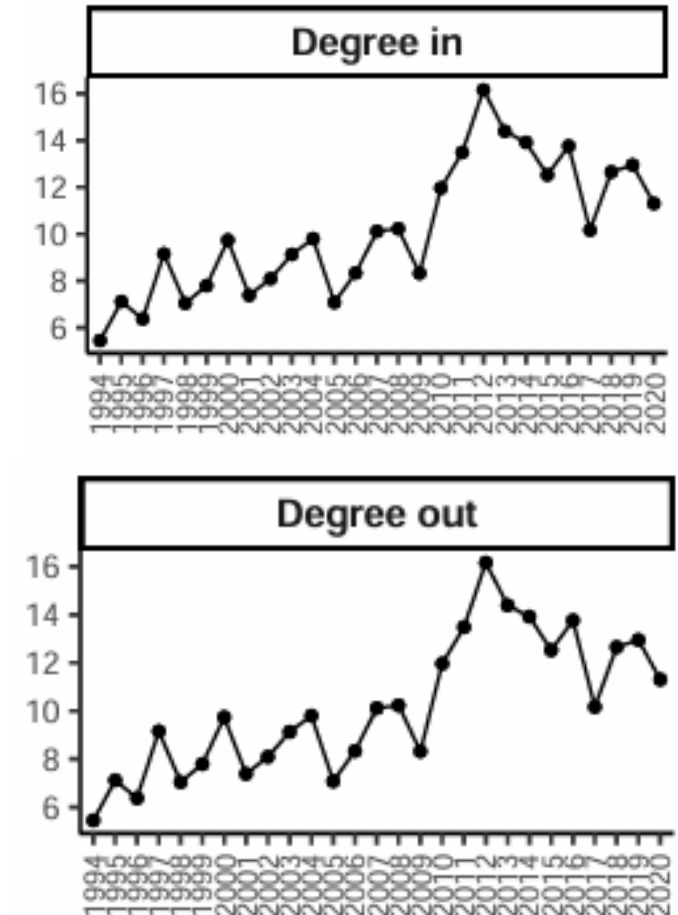
Progression of degree centrality during years

CENTRALITY

The drivers with the lowest betweenness (0 for almost every year) are low field drivers with some exception of top drivers.
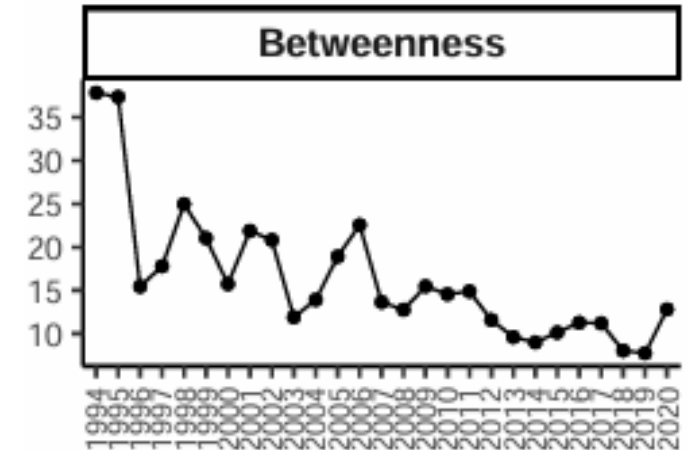The ones with highest betweenness are always drivers from the middle of the grid.

During the years this centrality has decreased, with some peak a few times.

Progression of betweenness centrality during years

The highest values of PageRank always correspond to mid field drivers. While the drivers with the lowest values some years are top drivers and some years are low field drivers.

This centrality, following the degree centrality, has increased during the years, but has not the flexion after the 2012.

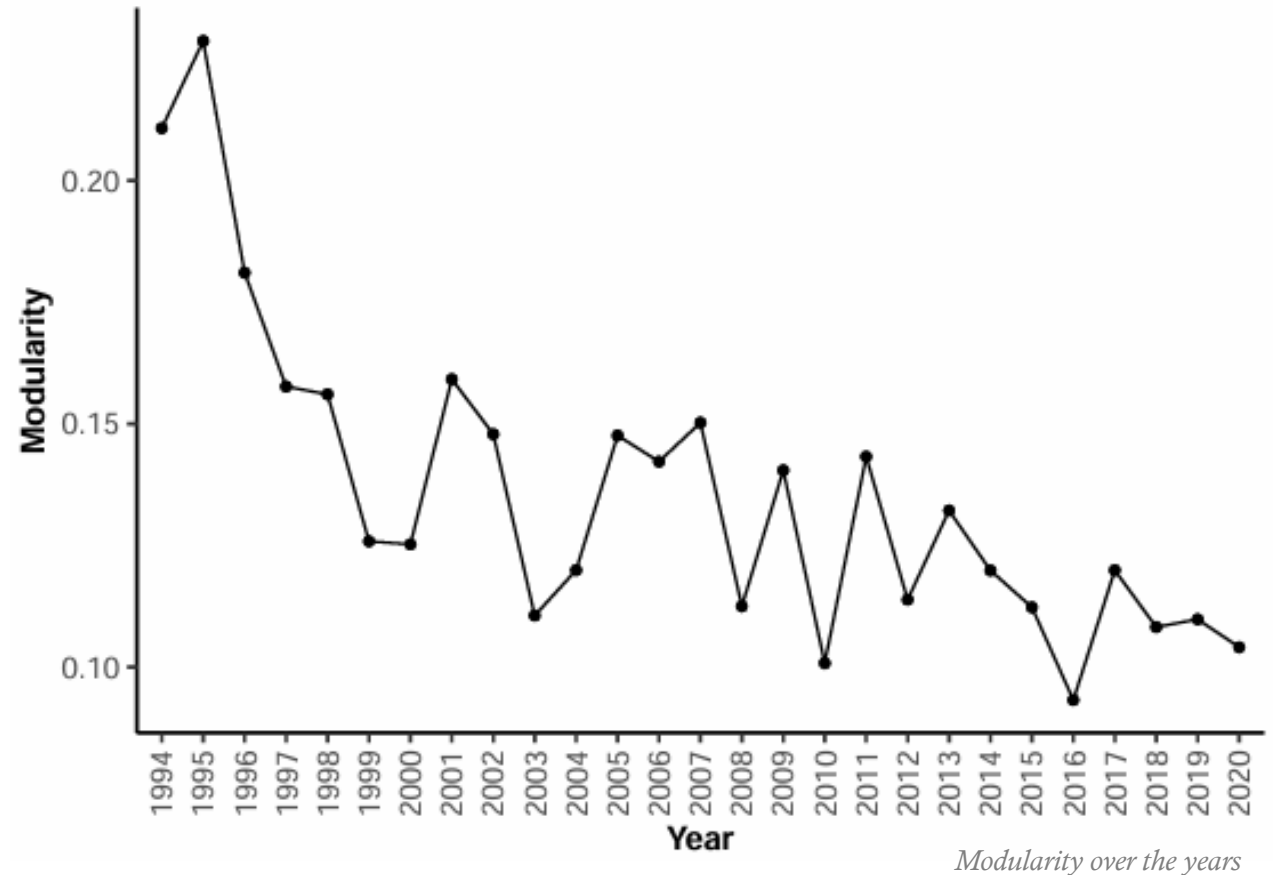*Progression of PageRank centrality during years*

MODULARITY

The modularity of the general graph is 0,1019.

The graphs have a positive value of modularity --> the nodes are organised in communities

The modularity value is pretty low --> the communities aren't definied

Over the years the modularity has a decreasing trend

Modularity over the years

CONNECTIVITY

The general graph has an high connectivity, near 95%

This time there is no trend in the annual graphs, but all of them have a high connectivity. Some of them higher than the general one, others lower.
Some of them are fully conntected.

1994	1995	1996	1997	1998	1999	2000	2001	2002
72,50%	94,29%	79,17%	100,0%	95,65%	100,0%	100,0%	92,31%	95,65%

2003	2004	2005	2006	2007	2008	2009	2010	2011
83,33%	92,00%	88,00%	96,30%	92,31%	100,0%	92,00%	92,59%	89,66%

| 2012	2013	2014	2015	2016	2017	2018	2019	2020
100,0%	100,0%	95,65%	100,0%	96,00%	83,33%	100,0%	100,0%	100,0%

General
93,79%

RESILIENCE

I chose a sample of graphs to analyze with different connectivity percentages (82,3%, 100%, 72,5%).
All graphs are pretty resilient, the best attacks are the ones that follow the degree and the betweenness centrality.
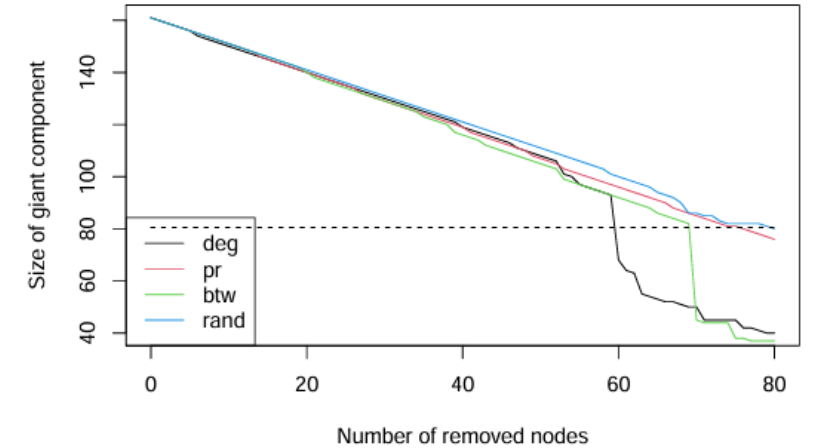The general graph has some key nodes which removal causes a drop, while in the others there is a linear reduction.
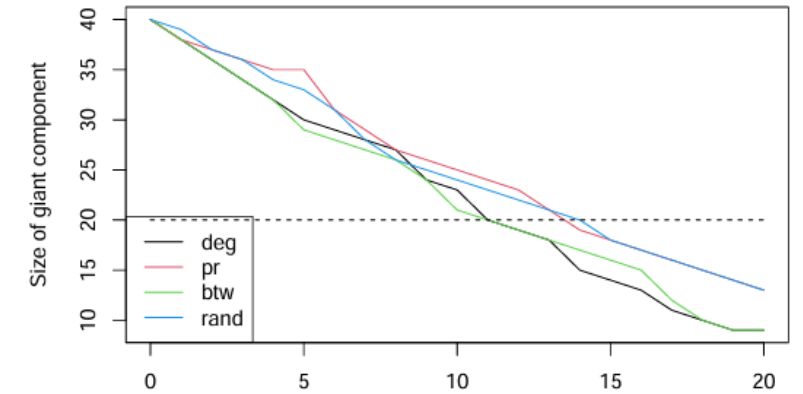

Percolation of the general graph
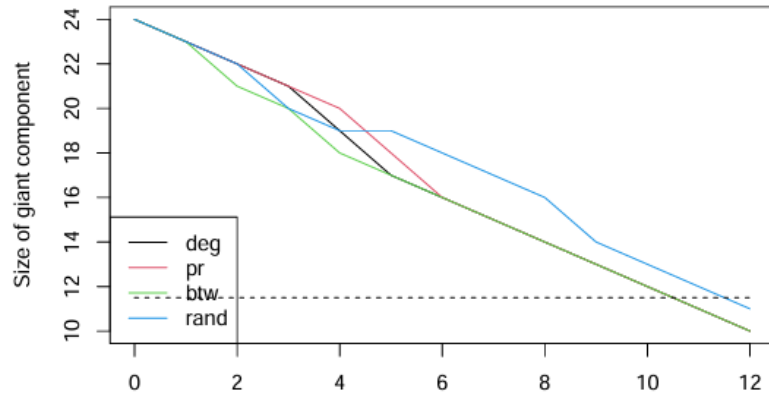
Percolation of the 2017 season
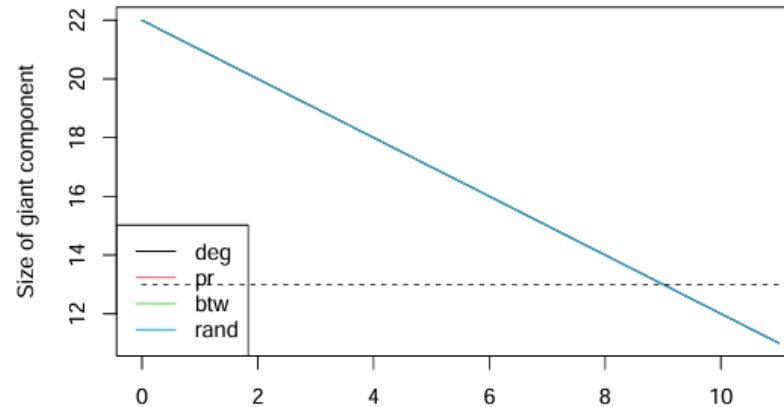
Percolation fo the 2008 season

Percolation of the 1994 season

DISTANCES AND DIAMETERS

General
3
11

All graphs (general and annuals) have short mean distances. The most surpising one is the general one that has far more nodes than the annual ones.

Also the diameters are very low.

With these data we can say that all the graphs are small world networks.

1994	1995	1996	1997	1998	1999	2000	2001	2002
2,32	2,16	1,83	1,71	2,19	1,91	1,72	1,95	1,99
5	5	4	3	5	5	5	5	6

2003	2004	2005	2006	2007	2008	2009	2010	2011
1,61	1,63	1,89	1,90	1,59	1,61	1,70	1,60	1,59
4	4	5	5	3	4	3	4	4

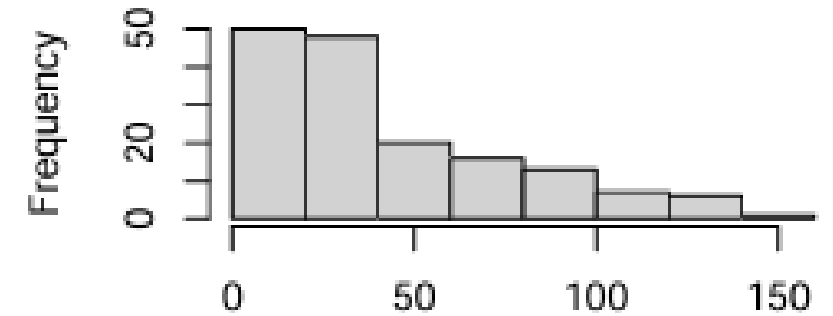| 2012	2013	2014	2015	2016	2017	2018	2019	2020
1,48	1,44	1,43	1,51	1,49	1,58	1,42	1,41	1,58
5	4	4	5	4	4	4	4	4

DEGREE DISTRIBUTION

The general graph has a degree distribution that seems to be a power law.
A lot of nodes with low degree and few nodes with high degree.

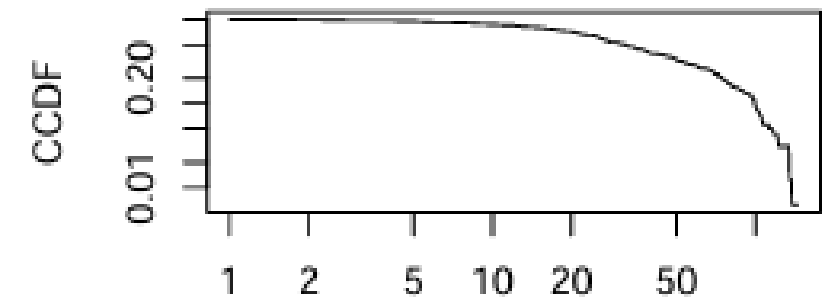*Degree distribution of the general graph*

The CCDF (Complementary Cumulative Distribution Function) on a log-log scale of a power law distribution is a straight line.
The one of the general graph is not a line, but for medium-high degree values is similar to a straight line.


CCDF of the general graph

DEGREE DISTRIBUTION

Even for the degree distribution I analyzed only a sample of the annual graphs, this time I took the first, the one in the middle and the last one.

In the graph of the 1994 season the distribution is similar to what we saw in the general graph, but less defined.
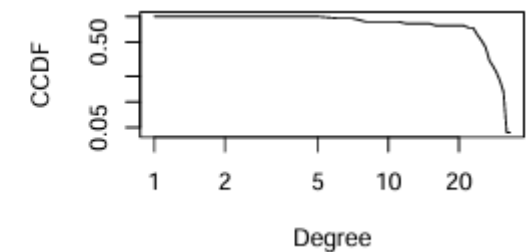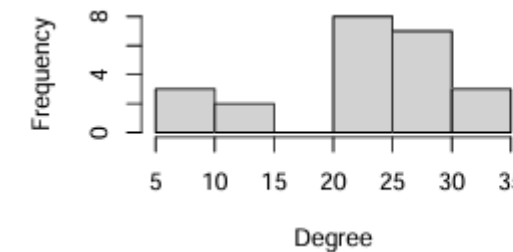
In the other 2 graphs for sure the distribution is not a power law, but it is more similar to a poisson distribution (more visible in the 2007 season).
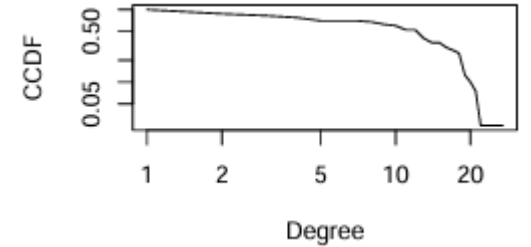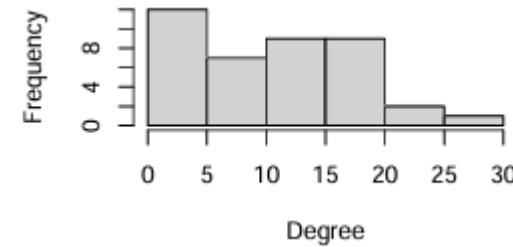
Degree distribution and CCDF of the 1994 season
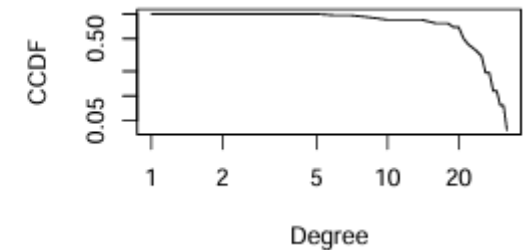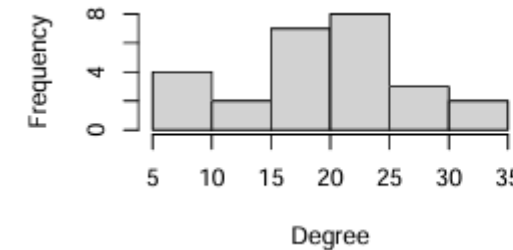
Degree distribution and CCDF of the 2007 season

Degree distribution and CCDF of the 2020 season