Podium and Influence: A Network Analysis of the Most Important Formula One Drivers

José Geraldo Rigotti de Faria[1], Fernando Guimarães Ferreira[2]

[1]Rigotti Consultoria - Avenida Paranaíba, 1598 - 77803-100 - Araguaína - TO - Brazil

[2]Instituto Infnet - Rua São José, 90 - 20010-020 - Rio de Janeiro - RJ - Brazil

rigotti@rigotticonsultoria.com.br, fernando.gferreira@infnet.edu.br

Abstract. *This paper presents a network analysis to recognize the most important Formula One drivers, focusing on podium finishes as a metric of influence. We investigated a dataset comprising 855 drivers, 1,079 races, and 25,840 results to explore the relationships and performance dynamics among drivers. The use of centrality measures and community detection in network analysis offers a novel approach to evaluating Formula One drivers, beyond traditional metrics like championship or race wins. In our comprehensive analysis, it was determined that PageRank yielded the most insightful results, effectively capturing the essential contributions and setting a new benchmark for assessing excellence in Formula One racing and other sports.*

1. Introduction

Formula One is nowadays a worldwide celebrated motorsport. This history began in 1950, with Nino Farina's conquer of the first championship driving an Alfa Romeo [Jones 2023]. Several legendary drivers and teams have succeeded in this highly competitive environment. Technological innovations, exceptional skills, and quick and strategic decisions differentiate the best in the Formula One world.

The current cars are very different from the starting years: faster, heavier and far more complex. Safety has been improved in the most recent decades. Despite these changes, the fans express their passion, the competition among teams and drivers remains intense, and the media's fascination with Formula One is remarkable. [Hamilton 2022]

And a question always arises: who are the most important drivers in the history of Formula One? Although simple, it provokes an intense debate among fans and motorsport experts. How should we measure importance?

In this paper, we propose a new approach to answering this intriguing question. The word "importance" here refers to evaluating the significance or influence of drivers within Formula One, assessed through their race outcomes and historical standing. Through the analysis of complex networks, we explored the connections and influences among drivers, based on an objective criterion: the drivers' presence on podiums and with whom the podiums were shared.

Earning a place on the podium is a significant triumph for a Formula One driver. This is a special recognition only granted by three drivers per race. Collecting multiple podium finishes, especially when shared with pilots who have achieved similar success, distinguishes the most competitive drivers.

While a simple count of podium finishes provides a basic success metric, it lacks depth in understanding the broader context and interactions within Formula One. Complex network analysis goes beyond mere frequency to explore how drivers' performances are interconnected through shared podiums, revealing patterns and influences not visible through simple statistics. This method can identify central figures within the sport's competitive network, providing a more comprehensive picture of a driver's influence and significance in Formula One.

The concept of this study was developed to assess the influence of each driver through centrality measures and community analysis. The main objective is to disclose, based on podium finishes, which drivers have the greatest importance in Formula One history, from 1950 to 2022, providing insights and contributing to the comprehension of the sport's structure and dynamics. We also aim to illustrate how complex network analysis can be effectively applied in sports analytics, offering a framework for evaluating athlete influence in other sports.

2. Related Work

Network analysis has emerged as a powerful tool in sports analytics, offering insights into patterns, relationships, and performance metrics that traditional analysis might overlook. The application of network analysis already extends to sports like tennis, rugby, soccer and basketball.

A dynamic network-based ranking system tailored to the nuanced performance shifts in competitive sports, notably applied to men's professional tennis, was demonstrated [Motegi and Masuda 2012]. Rugby team performance was analyzed using dynamic network analysis to predict outcomes by studying pass and disruption networks [Cintia et al. 2016]. An innovative study explored the application of network analysis combined with a statistical model to assess pass difficulty, thereby identifying key soccer players [McHale and Relton 2018]. Drawing on the analogy with search engine optimization techniques (PageRank), a study developed a network-driven approach for ranking and predicting outcomes in basketball leagues using directed graphs based on game results [Xia et al. 2018].

While the application of network analysis has shown promising results in various sports for understanding team dynamics, player performance, and predicting outcomes, its use in Formula One has been notably absent, particularly in the context of identifying the sport's most influential drivers.

A study on the competitive balance in Formula One was conducted, utilizing a simple econometric model, highlighting that regulation changes increased championship uncertainty without affecting race outcomes or dominance [Judde et al. 2013]. Another study introduced various measures and identified challenges related to data and methodologies [Budzinski and Feddersen 2019].

The escalation from competition to conflict was investigated, utilizing panel data on Formula One races from 1970 to 2014, focusing on the role of structural equivalence in competitive networks [Piezunka et al. 2018].

A novel Bayesian multilevel rank-ordered logit regression method to model individual race finishing positions was presented [Van Kesteren and Bergkamp 2023], allowing inferences about driver skill and constructor advantage.

Another study used group-based trajectory modelling to classify Formula One drivers from the turbo era [Piquero et al. 2021]. This article clearly stated that "there are few empirical studies that have developed and/or applied rigorous methodological techniques to examine which drivers are the most successful".

A social network analysis was utilized to investigate the economic dynamics of Formula One since the 2000 season, specifically focusing on the economic behavior of championship teams in their driver acquisitions [Mourao 2024]. This analysis examines the centrality of championship teams, their strategic selectiveness in driver movements, and the economic implications of their decisions.

These studies primarily focus on traditional statistical methods, performance metrics, and qualitative analyses to evaluate Formula One drivers. However, none of these studies employs network analysis to explore the importance of Formula One drivers. This absence presents a significant gap in the literature, suggesting an opportunity to apply this methodology to Formula One.

3. Methodology

This study aims to quantify the significance of Formula One drivers by employing network analysis and graph theory. For this reason, drivers are depicted as network nodes, with the podium finish serving as an element for establishing connections among drivers. In Formula One, a podium is shared by three drivers[1], thereby creating relationships that are represented as edges within the network. When two drivers share podium finishes on multiple occasions, the connection between them is assigned a greater weight.

Certain podium combinations serve as illustrations of different eras in Formula One, such as Hamilton, Bottas, and Verstappen; Mansell, Piquet, and Prost; Senna, Prost, and Berger; Schumacher, Raikkonen, and Barrichello. These examples of frequent trios likely constitute areas of strong connection within the constructed network, reflecting significant partnerships and rivalries that have shaped the sport's history [Kanal 2021]. These clusters highlight periods of dominance by certain drivers and suggest the evolving competitive dynamics and technological innovations within Formula One over the years.

3.1 The dataset

We used the dataset "Formula 1 World Championship (1950 - 2023)" [Vopani 2023, August]. The dataset had reliable data until the 2022 Championship.

[1] There were some exceptions, with podiums featuring 2, 4, 5 and 7 drivers, as explained in the section 4.1

3.2 The value of each race

The number of races per season has steadily grown throughout the studied period (Figure 1), varying from 7 races in 1950 to 22 races in 2021 and 2022 [Hamilton 2022].

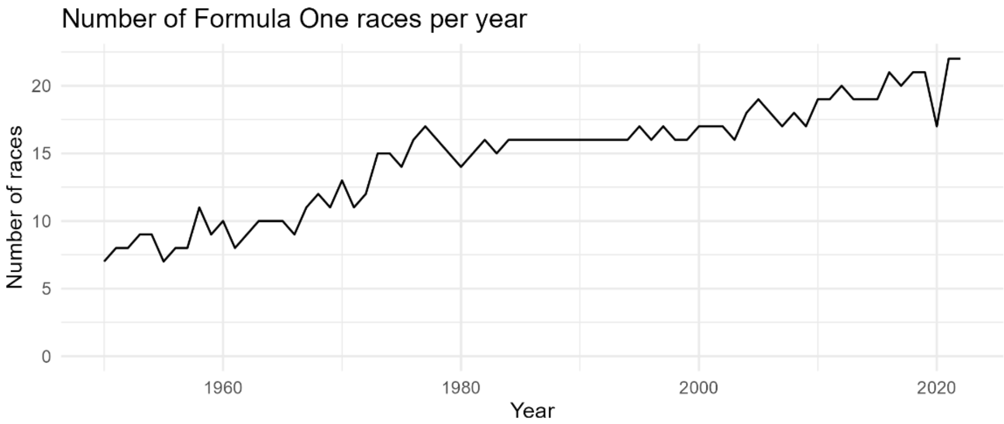


Figure 1. Number of Formula One races per year (1950-2022)

We considered that winning a race should have not the same value in our study. To achieve an unbiased value for each race in the dataset, an adjustment was applied to find a weight for each race using the following Equation 1:

$$Value\ of\ each\ race = \frac{Maximum\ number\ of\ races\ in\ a\ season}{Number\ of\ races\ in\ the\ selected\ season}\ (1).$$

So, a race in the 1950 season had a weighted value of 3.14 while a race in 2022 had a baseline value of 1.

3.3 Defining weighed values for the drivers' relations on the podium

In our network analysis, the edges, representing the competitive relationships and consequently the connectivity between drivers, are not uniformly weighted. This reflects the intuitive understanding that the rivalry and connection between drivers finishing in 1st and 2nd places are more pronounced than those between the 1st and 3rd places. To accurately quantify the interactions among drivers on the podium, we employed a weighted approach that is based on the points allocated to podium finishers. This methodology allows us to capture the varying degrees of competitiveness inherent in the finishing positions, with higher points indicating stronger connections.

Specifically, we applied values that correspond to the points system utilized in the Formula One World Championship, drawing upon the scoring regimes from different eras: 10, 8, and 6 points for podium finishes from 2003 to 2010, and 25, 18, and 15 points for podium finishes from the 2010 season onwards. So the edges were valued at 100% for the 1st and 2nd places, 80% for the 2nd and 3rd places and 60% for the 1st and 3rd places.

The resulting graph is undirected, indicating that the relationships between the drivers on the podium are bidirectional, reflecting mutual competitiveness irrespective of the direction of comparison.

3.4 Network Analysis

The next step involved examining the number of network components, specifically identifying if there were components apart from the giant component. The process then progressed to filtering, focusing solely on the giant component. It was imperative to ensure that any components excluded in this filtering process were not significant to the overall analysis. This step is crucial for concentrating the analysis on the most interconnected and relevant part of the network, which is typically represented by the giant component.

Furthermore, degree centrality, closeness, betweenness, weighted degree and PageRank [Ge and He 2022], were evaluated. Communities were identified using the Louvain method [De Meo et al. 2011]. Finally, the graph visualisation was performed, where the nodes were partitioned according to the modularity class.

3.5 The tools behind our data analysis

The dataset was processed using R version 4.3.2, adhering to the methodology outlined previously, to acquire the nodes and edges files. Subsequently, the analyses were performed utilizing Gephi 0.10. The technical tools and software versions employed in the study ensure transparency and reproducibility of the analysis process.

4. Results

This section presents a comprehensive analysis of the processed data and the resulting complex network. We examine the findings to elucidate the intricate dynamics of each driver’s importance, discussing the centrality measures and the modularity class.

4.1 Retrieved and filtered data

These are the figures for the retrieved data for the Formula One championships from 1950 to 2022: 855 drivers, 1,079 races and 25,840 results. After filtering the drivers to consider only those who achieved at least one podium, the number of drivers dropped to 215.

We had an unexpected finding: podiums with 2, 4, 5 or 7 drivers. The podium with 2 drivers happened in Brazil in 1983. The runner-up was disqualified and for the first and unique time in Formula One history the third and fourth-place finishers had their positions maintained and not upgraded [Motor Sport Magazine n.d.]. The podiums with 4, 5 or 7 drivers only happened in 18 races until 1960, especially in races in the United States, when more than one driver could share one car in the same race [Jones 2023].

4.2 Network evaluation

The nodes file had 215 drivers, while the edges file had 948 records after summing up the 3284 original edges.

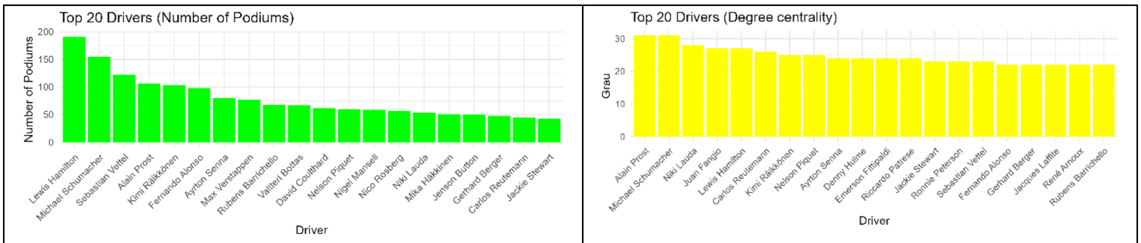
Four connected components were found. The three smaller components were formed by 25 American drivers who participated in a few races from 1950 to 1960, achieved a maximum of 4 podiums and had little historical importance [Jones 2023].

For this reason, and according to our methodology, we filtered only the giant component, which contained 190 drivers and 911 edges (more than 96% of the edges).

The five centrality measures explained below were used to assess the importance of the nodes in the Formula One network, each one offering a unique perspective. The degree centrality focuses on the number of edges (direct connections) of a node. The weighted degree centrality is similar to the degree centrality, enhancing the concept by considering the weight of each edge. The closeness centrality evaluates the average distance to other nodes. The betweenness centrality checks how often a node is on the shortest path between other nodes. The PageRank centrality, created to rank web pages, evaluates both the quantity and the quality of the node connections.

The top 20 drivers with the best numbers (described here as "Top 20") were selected for each measure. The number 20 represents approximately 10% of the nodes and is a cutoff point that can be considered fair throughout the studied period (73 years) since it allows a driver to stand out in that criterion every 3 to 4 years.

Number of podiums per driver. We expected a correlation between the number of podiums and the importance of the driver (Figure 2-A). The two major highlights here are Hamilton and Schumacher. All the first 8 drivers are or were World Champions. A very important remark: only 5 among the Top 20 drivers were not World Champions: Barrichello, Bottas, Coulthard, Berger, and Reutemann. However, these numbers are raw and unweighted. Fangio does not appear on the list, and Stewart is only the twentieth. Therefore, the criterion involving a pure number of podiums is unfair to the drivers who competed in the early decades of Formula One when there were fewer races per season.



(A) (B)

Figure 2. Top 20 drivers according to number of podiums (A) and degree centrality (B)

Degree centrality. We are essentially valuing the number of connections among the nodes (Figure 2-B). The average degree centrality is 9.6 (ranging from 2 to 31). The degree distribution has a long-tail aspect. The top 5 drivers are very well-known World Champions: Prost, Schumacher, Fangio, Lauda, and Hamilton. Among the next 15 ranked, 7 were not World Champions: Reutemann, Patrese, Peterson, Berger, Laffite, Arnoux, and Barrichello. It's worth noting that all these were very competitive drivers for a significant part of their careers. But it's odd, for example, that the list does not contain the three-time champion Jack Brabham. It also raises the question: how does Prost, with fewer podiums than Hamilton or Schumacher, surpass the two? Or how to explain Fangio ahead of Hamilton? Or even Reutemann in sixth? The answer lies in speculation, which requires a complementary study. It is quite reasonable to assume that the variation in podium compositions was greater when Prost and Reutemann were racing. There was not an absolute dominance by one team as is currently the case with Red Bull and was recently with Mercedes. And if the competition for the first place is greater, most likely the same happens with the other two podium positions, generating greater diversity in the compositions. That is, the degrees of Prost and Fangio may have a lot of correlation with this greater plurality. In summary, although degree centrality is associated with some significant distortions, it represents a great advancement over the simple counting of races or podiums.

Weighted degree centrality. This centrality is associated with a crucial point of the paper: the weight of the edges. According to this reasoning, the weight of the edges would have greater importance than the number of connections of each node (Figure 3-A). The two first drivers, by a notable lead ahead of the others, are Hamilton and Schumacher. In this case, the number of podiums probably made a lot of difference, as each podium directly impacts the weight of the edges. Prost is third in this criterion, surpassing Vettel when compared to degree centrality. The diversity referred to in the previous analysis was probably the cause of this inversion. The top eight were or are World Champions. Of the remaining 12, only 3 never won a title: Barrichello, Coulthard, and Bottas. Considering that a world championship title is a privilege of few drivers, we see here an advancement over degree centrality, as, in the latter, 7 have never been champions. Again, compared to degree centrality, Reutemann, Patrese, Fittipaldi, Hulme, Petterson, Arnoux, Berger, and Laffite left the list. Verstappen, Coulthard, Mansell, Graham Hill, Bottas, Clark, Hakkinen, and Brabham entered. 3 world championships left. 12 championships entered. Now comparing with the simple counting of the number of podiums, the non-champions Berger and Reutemann left the list. Champions Button and Nico Rosberg also left. In return, Fangio, Graham Hill, Clark, and Brabham entered. There is a huge difference from the point of view of the number of world championships. 2 titles left. 12 entered. Thinking qualitatively, this criterion is much better than the simple count of podiums and also clearly superior to degree centrality.

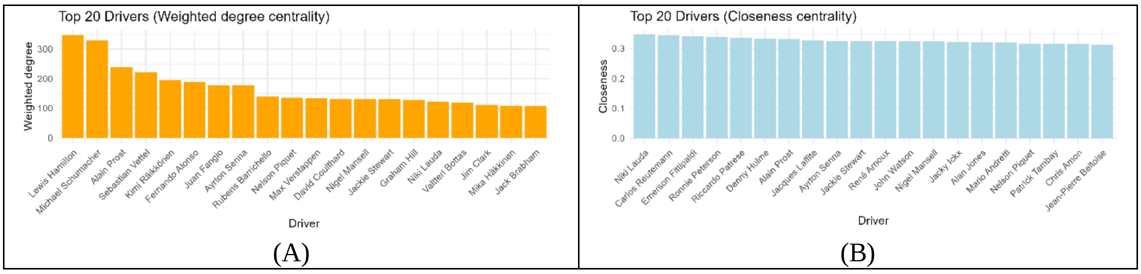


Figure 3. Top 20 drivers according to weighted degree centrality (A) and closeness centrality (B)

Closeness centrality. (Figure 3-B). This criterion has a huge weakness in this specific network. The distribution of the drivers occurs over just over 70 years. It is natural that those who raced in a middle time range, especially in the 1980s, have a clear advantage over all others, which is undoubtedly expressed in the figure. Indeed, all the drivers among the Top 20 competed in races in the 1970s and, especially, in the 1980s. Such a fact renders this criterion useless for this paper.

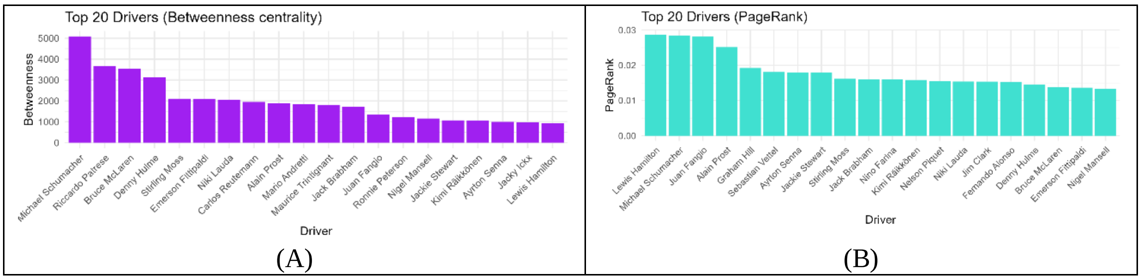


Figure 4. Top 20 drivers according to betweenness centrality (A) and PageRank (B)

Betweenness centrality. (Figure 4-A). For this criterion, we also have a natural disadvantage for the drivers who are at the extremes of the time range, although not as significant as in the previous item. The great highlight goes to Schumacher, with a long career close to the central range of the considered period. Occupying from second to fifth position are Patrese, McLaren, Hulme, and Moss. All of them had long careers, especially Patrese, and, except for Moss, were also positioned close to the central range of the considered period. It is important to note that Hamilton is only twentieth in this criterion. Probably, in 20 years, Hamilton's position will rise significantly, as his career is quite long and he will no longer be positioned at an extremity of the period. In summary, this criterion is interesting for some networks but subject to relevant distortions and of little use in the context of this analysis.

PageRank. (Figure 4-B). It is crucial to remember that the weight of the edges, an essential part of this article, was considered in the computation of the PageRank. Hamilton, Schumacher, and Fangio stood out. Fangio received a fair evaluation, as he was a five-time champion. Among the Top 20 drivers, only 2 drivers, Moss and McLaren, were never World Champions. The Top 14 drivers included all 10 drivers with three or more world titles. Verstappen is already a three-time champion, but data from the 2023 season were not used. The 4 drivers who "intrude" into this select group are two-time champion Graham Hill, champions Farina and Raikkonen, and Moss. Hill is not an intruder, at most a positional inversion. Farina and Raikkonen were very important drivers in their respective eras. Of the 7 two-time champions, only Verstappen, Ascari, and Hakkinen are not on the list. Verstappen would probably be there if the 2023 season had been computed. The only real surprises among the Top 20 would be Moss and McLaren. Moss was runner-up for 4 consecutive years (from 1955 to 1958), three of those times to Fangio. In this sense, Moss was unlucky. McLaren indeed causes surprise. Throughout his career, he won only 4 races and was third in the world championship twice. A career overshadowed when compared to Verstappen, Ascari, or Hakkinen. However, McLaren was contemporary with many great names on the list: Graham Hill, Stewart, Moss, Brabham, Clark, Hulme, and Fittipaldi. These relationships were most likely the determining factor for the algorithm to include him in the Top 20 drivers. From the qualitative perspective of this analysis, McLaren would be the only questionable name from this Top 20. And Verstappen the main questionable absence, although reiterating that the 2023 season was not computed. All these observations already demonstrate how well PageRank managed to capture the relevance of the drivers. Therefore, the comparison of PageRank will be made directly with the centrality by degree weighted by the edges, which until then was the best scenario. Leaving the Top 20 list were

Barrichello, Verstappen, Coulthard, Bottas, and Hakkinen. Entering the PageRank Top 20 drivers were Moss, Farina, Hulme, McLaren, and Fittipaldi. 4 titles left from the two-time champions Verstappen and Hakkinen. And 4 titles entered from the two-time champion Fittipaldi and champions Farina and Hulme. What is observed is almost a generational exchange. All 5 excluded drivers started racing in Formula One after 1990. All 5 included drivers had ended their activities by 1980. The change is fair, as there needs to be a balance between representatives of various generations. With degree centrality, only Fangio had ended his career by 1970. With the degrees weighted by the edges, the number rose to 4, still low. Finally, with PageRank, the number rises to 6, which is fairer.

Modularity class. We discovered five distinct communities within the sport, classified by modularity class. This finding underscores the presence of clusters within Formula One's extensive history, each representing a unique era of competition among drivers, encapsulating rivalries that have defined decades of racing.

We focused our analysis on the top 20 drivers identified by PageRank, as outlined in Table 1, detailing their rankings and the number of World Championships each has won. A notable discovery from this examination was that PageRank effectively highlighted the five most significant drivers in Formula One history, with each of these leading drivers being representative of a distinct community. This finding underscores PageRank's utility in pinpointing key figures within the sport's complex network of relationships and achievements.

Another remarkable observation is that Schumacher monopolizes the 4th community, whereas the 2nd community is extremely competitive. With seven members, McLaren seems to be an outlier, as stated before.

Table 1. Top 20 Formula One drivers' distribution across communities, ranked by PageRank (in parenthesis). The number of World Championships (WC) won is indicated within brackets.

1st Community (9 WC)	2nd Community (16 WC)	3rd Community (19 WC)	4th Community (11 WC)	5th Community (18 WC)
Juan Fangio (3) [5]	Graham Hill (5) [2]	Alain Prost (4) [4]	Michael Schumacher (2) [7]	Lewis Hamilton (1) [7]
Stirling Moss (9) [-]	Jackie Stewart (8) [3]	Ayrton Senna (7) [3]		Sebastian Vettel (6) [4]
Nino Farina (11) [1]	Jack Brabham (10) [3]	Nelson Piquet (13) [3]		Kimi Räikönnen (12) [1]
	Jim Clark (15) [2]	Niki Lauda (14) [3]		Fernando Alonso (16) [2]
	Denny Hulme (17) [1]	Nigel Mansell (20) [1]		
	Bruce McLaren (18) [-]			
	Emerson Fittipaldi (19) [2]			

The next 5 drivers in the PageRank classification are shown in Table 2. Ascari, a two-time champion, would be a good replacement for McLaren. And if the 2023 data were considered this position probably would be of Verstappen, since he was the champion that year. Two drivers who were never champions appear: Reutemann and Barrichello. Reutemann was the runner-up in 1981 and the 3rd for three times. Barrichello was the runner-up two times and once finished 3rd. So it is clear that these two drivers were not included in the Top 25 by chance. Maybe being the runner-up is somehow undervalued and Moss is a good example of this situation.

Table 2. Formula One drivers' (from rank 21 to 25) distribution across communities, ranked by PageRank (in parenthesis). The number of World Championships (WC) won is indicated within brackets.

1st Community (9 WC)	2nd Community (16 WC)	3rd Community (19 WC)	4th Community (11 WC)	5th Community (18 WC)
Alberto Ascari (21) [2]	John Surtess (23) [1]	Carlos Reutemann (22) [-]	Rubens Barrichello (24) [-]	Max Verstappen (25) [2]

An additional analysis presented in Table 3 focuses on World Champion drivers who are ranked beyond the top 25 in the PageRank classification. Among these, Häkkinen stands out as the sole two-time champion. Alongside him, twelve drivers have each conquered the championship once. Given this context, Häkkinen could arguably be considered for inclusion within the top 20 drivers due to his multiple championships. However, justifying the inclusion of the other champions, who have won the title only once, into the top 20 proves to be more challenging.

Table 3. Formula One World Champion drivers' (not included in the Top 20 ranking) distribution across communities, ranked by PageRank (in parenthesis). The number of World Championships (WC) won is indicated within brackets.

1st Community (9 WC)	2nd Community (16 WC)	3rd Community (19 WC)	4th Community (11 WC)	5th Community (18 WC)
Mike Hawthorn (26) [1]	Phil Hill (38) [1]	Jody Scheckter (33) [1]	Mika Häkkinen (34) [2]	Nico Rosberg (41) [1]
	Jochen Rindt (55) [1]	James Hunt (46) [1]	Damon Hill (35) [1]	Jenson Button (43) [1]
		Alan Jones (47) [1]	Jacques Villeneuve (64) [1]	
		Mario Andretti (61) [1]		
		Keke Rosberg (66) [1]		

For illustrative purposes, Figure 5, generated with Gephi, displays a portion of the network, specifically the 5th community and the end part of the 4th community. This is an example to provide insight into the network's structure, showcasing the intricate connections and structure within the network.

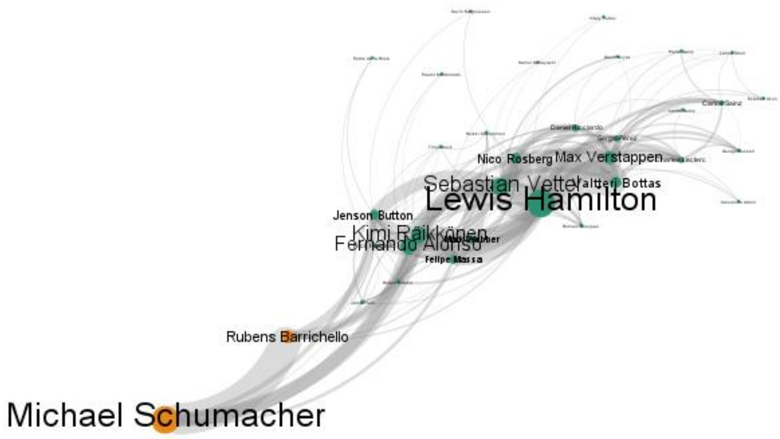


Figure 5. Partial Network Visualization showcasing the 5th Community and the final segment of the 4th Community, generated with Gephi

In generating the visualization, the Force Atlas 2 layout was utilized, incorporating various adjustments to optimize the network's clarity and readability: 'discourage hubs' (to reduce the prominence of central nodes), 'prevent overlap' (to eliminate overlapping elements for a cleaner visual) and the most effective scaling and gravity settings.

This synthesis of historical data and advanced network analysis, particularly the insightful use of PageRank, highlights the most important Formula One drivers.

5. Limitations and Future Work

In this section, we examine the constraints of our current methodology and outline the opportunities for enriching our understanding of influence within Formula One racing and other sports.

5.1 Limitations

Focusing on podium finishes as the primary measure of success and connectivity and on the top 20 drivers might not capture the complete spectrum of influence. Expanding the criterion and including the top six finishers and/or including more drivers, for instance, could potentially highlight the contributions of drivers who consistently perform well but may not always secure a podium position.

Furthermore, altering the specific weighted values assigned to the relationships between drivers might impact the perceived influence and centrality of certain drivers within the network.

Additionally, the nature of our analysis is drawn from historical data up to the 2022 season, reflecting past and present dynamics without forecasting future developments.

5.2 Future work

Looking ahead, several avenues for future research emerge from this study. Incorporating and weighting additional performance metrics, such as qualifying positions and fastest laps, could provide a more nuanced understanding of a driver's influence.

In addition, further studies could also include weighing the performance of constructors in Formula One, giving additional weight to podium finishes achieved by drivers who are not in the leading constructors' cars.

Moreover, updating the analysis to include data from the most recent Formula One seasons (2023 and forward) would ensure that our findings remain relevant.

Finally, the methodologies applied in this study hold promise for broader applications beyond Formula One. Similar network analyses could be conducted within other sports contexts, particularly in individual sports like tennis, golf, boxing, and swimming. These sports, characterized by rich histories of rivalry, strategy, and performance, present fertile ground for exploring the applicability of network analysis in capturing the essence of athletic influence and legacy.

6. Conclusion

Our study offers a unique perspective focused on podium finishes as a metric of success, analyzing the connectivity among Formula One drivers from 1950 to 2022. By leveraging complex network analysis to investigate the podium-based influences among drivers, we shed new light on who the key drivers in Formula One history are.

This approach allowed us to transcend traditional statistics, providing a deeper understanding of the dynamics and interrelations that have shaped the sport over the decades and demonstrating the powerful application of data science in sports.

Our main finding was obtained using PageRank centrality and showed Hamilton, Michael Schumacher, Fangio, Prost and Graham Hill as the most important drivers. Each one of them represents a community in modularity class, thus providing a balanced view of the sport.

Our results validated the methodology and revealed insights into the factors that define the drivers' impact in Formula One racing.

To support our findings, the data and code used in this study are available for review and replication. They can be accessed on GitHub here: https://github.com/josegeraldorigotti/brasnam2024-f1-network-analysis. This open access fosters further research and discussion within the sports analytics community.

We acknowledge limitations in using podium finishes as a sole success metric and focus on the top 20 drivers. Our proposal is to refine the analysis with additional performance metrics such as qualifying positions and fastest laps.

This study's insights and methodologies enhance the understanding of Formula One driver's influence and create a framework for deeper exploring complex networks in a wide variety of sports, like tennis or golf.

References

Budzinski, O., and Feddersen, A. (2019). Measuring Competitive Balance in Formula One Racing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3357687

Cintia, P., Coscia, M., and Pappalardo, L. (2016). The Haka network: Evaluating rugby team performance with dynamic graph analysis. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1095-1102). IEEE. https://doi.org/10.1109/ASONAM.2016.7752377

De Meo, P., Ferrara, E., Fiumara, G., and Provetti, A. (2011). Generalized Louvain method for community detection in large networks. In *2011 11th International Conference on Intelligent Systems Design and Applications (ISDA)* (pp. 88-93). IEEE. https://doi.org/10.1109/ISDA.2011.6121636

Ge, Y., and He, Y. (2022). Research on Importance and Correlation Evaluation of Music Network Nodes Based on Page Rank Algorithm. *Journal of Applied Mathematics and Computation*, 6(2), 267-272. DOI: 10.26855/jamc.2022.06.012

Hamilton, M. (2022). *Formula 1: The Official History.* Welbeck.

Jones, B. (2023). *The Formula One Record Book: Grand Prix Results, Team & Driver Stats, All-Time Records.* Welbeck. Kindle Edition.

Judde, C., Booth, R., and Brooks, R. (2013). Second Place Is First of the Losers: An Analysis of Competitive Balance in Formula One. *Journal of Sports Economics*, 14(4), 411-439. https://doi.org/10.1177/1527002513496009

Kanal, S. (2021, May 29). After Bottas, Hamilton and Verstappen, which other trios have shared an F1 podium most often? *Formula 1*. https://www.formula1.com/en/latest/article.after-bottas-hamilton-and-verstappen-which-other-trios-have-shared-an-f1.2ClY69iuDoyIW4rFrdzPTe. Accessed: 2024_03_22

McHale, G. and Relton, S. D. (2018). Identifying key players in soccer teams using network analysis and pass difficulty. *European Journal of Operational Research*, 268(1), 339-347. https://doi.org/10.1016/j.ejor.2018.01.018

Motegi, S. and Masuda, N. (2012). A network-based dynamical ranking system for competitive sports. *Scientific Reports*, 2, 904. https://doi.org/10.1038/srep00904

Motor Sport Magazine. (n.d.). 1983 Brazilian Grand Prix. Retrieved from https://www.motorsportmagazine.com/database/races/1983-brazilian-grand-prix/. Accessed: 2024_03_22

Mourao, P. R. (2024). Drivers' moves in Formula One Economics: A network analysis since 2000. *SPORT TK-Revista EuroAmericana de Ciencias del Deporte*, 13, 9. https://doi.org/10.6018/sportk.540441

Piezunka, H., Lee, W., Haynes, R., and Bothner, M. S. (2018). Escalation of competition into conflict in competitive networks of Formula One drivers. *Proceedings of the*

National Academy of Sciences of the United States of America, 115(15), E3361-E3367. https://doi.org/10.1073/pnas.1717303115

Piquero, A. R., Piquero, N. L., and Han, S. (2021). Identifying the Most Successful Formula 1 Drivers in the Turbo Era. *The Open Sports Sciences Journal*, 14, 151-157. https://opensportssciencesjournal.com/VOLUME/14/PAGE/151/

Van Kesteren, E.-J. and Bergkamp, T. (2023). Bayesian analysis of Formula One race results: Disentangling driver skill and constructor advantage. *Journal of Quantitative Analysis in Sports*, 19(4), 273–293. https://doi.org/10.1515/jqas-2022-0021

Vopani. (2023, August). Formula 1 World Championship (1950 - 2023). Kaggle. https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020/. Accessed: 2024_03_22

Xia, V., Jain, K., Krishna, A., and Brinton, C. G. (2018). A network-driven methodology for sports ranking and prediction. In *2018 52nd Annual Conference on Information Sciences and Systems (CISS)* (pp. 1-6). IEEE. https://doi.org/10.1109/CISS.2018.8362324