Analyzing Opening Choices and Player Strengths in Online Chess: A Network Analysis Approach

Taner Dinçer, Serkan Kütük, Gürkan Talha Soylu, Özge Karasu, Uğur Öztunç

Abstract

Chess, a widely enjoyed board game played by millions of people, has seen a significant increase in popularity, particularly during the pandemic. Despite varying levels of skill and performance, understanding the differences in player strengths remains an interesting area of study. The opening moves of a game play a crucial role in its outcome, making the investigation of opening choices important. In this paper, we analyze opening preferences among players of different skill levels using network analysis techniques. By examining data from an online chess league, we construct a player-opening bipartite network to explore the relationship between openings and player strengths. Additionally, we compare players of different skill levels using community detection methods. Our results reveal intriguing and non-linear patterns between skill levels and opening choices. This study is very important in terms of chess literature, because after all, a correct opening choice might win you the game.

Keywords: Chess, Network Theory, Opening, Elo, Player strength, Network, Lichess4545, Online chess

1. Introduction

1.1 Background and Significance

Chess, a strategic board game played by millions of enthusiasts worldwide, has experienced a surge in popularity in recent years, with a notable increase during the global pandemic. As players engage in online platforms, massive amounts of game data become available for analysis, offering a unique opportunity to explore various aspects of chess gameplay. One intriguing area of study is understanding the differences in player strengths and how they relate to the choices made during the opening phase of a game. The opening moves in chess set the stage for the subsequent development and outcome of the game, making an investigation of opening choices crucial.

Traditionally, chess opening analysis has relied on the expertise of grandmasters and the examination of historical games. However, with the advent of online chess platforms and the availability of vast amounts of game data, researchers can now employ quantitative methods to analyze player behaviors and identify patterns. Network analysis, a powerful tool for studying complex systems, provides a valuable framework for examining the relationship between opening moves and player strengths in chess.

In this research paper, a network analysis approach is employed to investigate opening preferences among players of different skill levels using data obtained from an online chess league. By constructing a player-opening bipartite network, the interplay between players and their preferred opening moves is captured. This network representation allows researchers to explore the structure and dynamics of opening choices in relation to player strengths, providing insights into strategic preferences and potential advantages at different skill levels.

First, a simple network analysis by evaluating the centrality metrics of our network is conducted. After that, the relationship between pre-grouped openings and players from different skill levels is investigated. These analyses allow researchers to show what people prefer which openings from different skill levels. On top of that, community detection methods for the data, which enable us to identify groups of players with similar opening preferences and performances are employed. By analyzing these communities, it is aimed to uncover patterns that reveal distinctive strategies employed by players of varying skill levels. Furthermore, Elo ratings, a widely recognized measure of player strength in chess, to quantify and validate the skill differences observed in our analysis are utilized.

This research uncovers intriguing and non-linear relationships between player strengths and opening choices. It is observed that certain opening moves are more prevalent among players of specific skill levels, suggesting a link between strategic choices and player expertise. By examining the data from the online chess league Lichess4545, the existing chess literature by providing empirical evidence of the association between opening choices and player strengths is contributed.

The insights gained from this study have practical implications for both chess players and game analysts. Understanding the strategic preferences of players at different skill levels can inform training strategies and guide players in selecting openings that align with their strengths. Additionally, the findings of this study contribute to the broader field of network theory by showcasing its applicability in analyzing complex systems such as online chess.

1.2 The Dataset

The dataset used in the project is obtained from API of Lichess.org, which is one of the most popular chess platforms in the world ("About lichess.org," n.d.). Lichess, which includes players who enjoy playing chess games with long-term time control, provides the opportunity to compete in the Lichess4545 league for its users who enjoy such chess competitions ("About lichess.org," n.d.).

The dataset contains information on all chess competitions played in the Lichess4545 league between 2015.11.01 and 2023.01.23. In the dataset, where there are 31188 matches in total, the Elo-ratings of the players also vary considerably. The player with the lowest rating has an Elo score of 802, while the player with the highest rating has a score of 2641. There are 473 different opening types in the data set with 3498 players in total. In terms of handling such a wide scale, it is thought to be a very suitable dataset for the project's description and objectives.

Each game, in the league, is stored in the same format, which can be seen in Figure 1. For each record, the player names of the white and black players, the result of the game, the ECO type, and the specific opening name can be retrieved from the dataset.

```
[Event "Lichess4545 League Season 33"]
[Site "https://lichess.org/gWuHPZFo"]
[Date "2023.01.23"]
[Round "8"]
[White "pafiedor"]
[Black "glbert"]
[Result "1-0"]
[WhiteElo "1857"]
[BlackElo "1895"]
[ECO "D11"]
[Opening "Slav Defense: Modern Line"]
[TimeControl "2700+45"]
[UTCDate "2023.01.23"]
[UTCTime "20:00:35"]
[Variant "Standard"]
[Termination "Normal"]
[WhiteRatingDiff "+27"]
[BlackRatingDiff "-17"]

1. d4 { [%clk 0:45:00] } 1... d5 { [%clk 0:45:00] } 2. c4 { [%clk 0:45:21]
} 2... c6 { [%clk 0:45:37] } 3. Nf3 { [%clk 0:45:16] } 3... Nf6 { [%clk
0:46:18] } 4. g3 { [%clk 0:45:02] } 4... dxc4 { [%clk 0:46:28] } 5. Bg2 {
[%clk 0:45:28] } 5... Bf5 { [%clk 0:46:01] } 6. Nbd2 { [%clk 0:45:39] }
6... e6 { [%clk 0:44:05] } 7. Nxc4 { [%clk 0:46:12] } 7... Be7 { [%clk
0:43:57] } 8. O-O { [%clk 0:46:38] } 8... O-O { [%clk 0:44:39] } 9. Bf4 {
[%clk 0:44:18] } 9... Nd5 { [%clk 0:43:00] } 10. Bd2 { [%clk 0:44:55] }
10... Nd7 { [%clk 0:39:54] } 11. Nfe5 { [%clk 0:42:54] } 11... Nxe5 { [%clk
0:38:37] } 12. Nxe5 { [%clk 0:43:35] } 12... Nf6 { [%clk 0:38:26] } 13. Bc3
{ [%clk 0:40:22] } 13... Be4 { [%clk 0:35:58] } 14. f3 { [%clk 0:38:12] }
14... Bf5 { [%clk 0:35:03] } 15. e4 { [%clk 0:38:44] } 15... Bg6 { [%clk
0:35:25] } 16. h4 { [%clk 0:35:12] } 16... h6 { [%clk 0:33:40] } 17. Nxg6 {
[%clk 0:33:21] } 17... fxg6 { [%clk 0:34:21] } 18. Qb3 { [%clk 0:34:03] }
18... Qd7 { [%clk 0:34:24] } 19. Bh3 { [%clk 0:33:27] } 19... Kh7 { [%clk
0:34:28] } 20. Qxe6 { [%clk 0:32:04] } 20... Qc7 { [%clk 0:33:59] } 21. e5
{ [%clk 0:31:36] } 21... Nd5 { [%clk 0:34:13] } 22. h5 { [%clk 0:31:08] }
22... gxh5 { [%clk 0:32:55] } 23. Bf5+ { [%clk 0:31:47] } 23... Rxf5 {
[%clk 0:33:19] } 24. Qxf5+ { [%clk 0:32:27] } 24... Kh8 { [%clk 0:32:52] }
25. Qxh5 { [%clk 0:30:47] } 25... Rf8 { [%clk 0:32:58] } 26. Bd2 { [%clk
0:28:29] } 26... Qb6 { [%clk 0:32:16] } 27. Bxh6 { [%clk 0:27:13] } 27...
Qxd4+ { [%clk 0:32:17] } 28. Be3+ { [%clk 0:27:26] } 28... Kg8 { [%clk
0:33:00] } 29. Bxd4 { [%clk 0:28:10] } 29... b6 { [%clk 0:33:13] } 30. Kf2
{ [%clk 0:26:59] } 30... c5 { [%clk 0:33:47] } 31. Rh1 { [%clk 0:27:15] }
31... Rxf3+ { [%clk 0:34:04] } 32. Kxf3 { [%clk 0:27:39] } 32... Nf6 {
[%clk 0:34:37] } 33. exf6 { [%clk 0:27:46] } 33... Bxf6 { [%clk 0:35:17] }
34. Bxf6 { [%clk 0:28:27] } 34... gxf6 { [%clk 0:36:02] } 35. Rae1 { [%clk
0:28:34] } 35... Kg7 { [%clk 0:36:39] } 36. Re7+ { [%clk 0:28:27] } 36...
Kf8 { [%clk 0:37:22] } 37. Qf7# { [%clk 0:28:25] } 1-0
```

FIGURE 1

1.3 Objectives

The main objectives of the research:

- Investigating the relationship between the openings and the chess level of players, which are determined by their Elo-ratings, using visual network analysis and statistics

- Analyzing the chess community and different playing habits, by benefiting from network analysis techniques such as closeness and betweenness centrality, in the framework of network science

- Establishing a link between player level and gameplay habits using the Community Detection algorithm, Louvain method

1.4 Methodology

The subsequent data set invites a thorough examination into the complex realm of chess, embodying elements of strategy and intellectual prowess. It comprises a compilation of 3,498 unique chess enthusiasts, each distinguished by individual skill sets, playing styles, and varied levels of experience, unified by their shared appreciation for this historic game.

The provided data set boasts a strategic diversity that is unparalleled, encompassing a total of 473 distinct openings. This collection illuminates a captivating spectrum of styles and tactics, each opening acting as a delicate choreography of moves that lay the groundwork for the ensuing match. These openings are characterized by their ECO (Encyclopedia of Chess Openings) types, systematically arranged into five overarching categories: A, B, C, D, and E.

The crux of this data set resides in the Elo ratings, serving as numerical reflections of each player's skill level and experience. The Elo ratings encapsulated within this dataset extend from the modest inception at 806 to the formidable pinnacle of 2641. The straightforward Elo rating system facilitates the classification of players into distinct skill brackets: the dataset includes 40 burgeoning talents with an Elo range of 800-1200, 692 players refining their abilities within the 1201-1600 range, a substantial group of 2,090 players exhibiting proficiency within the 1601-2000 range, and 670 advanced tacticians boasting an impressive Elo rating of 2001-2400.

This comprehensive dataset unveils a multitude of analytical possibilities, ranging from exploring the popularity of different openings, and investigating the impact of player ratings on game outcomes, to the prediction of match results. The stage is set, offering an invitation to delve into the captivating world of strategic maneuvers and intellectual engagements.

2 Literature Review

2.1 Quantifying The Complexity And Similarity Of Chess Openings Using Online Chess Community Data

The relationship between chess players and the openings they use has always been a big question mark within the chess community. Topics such as the playing styles of great chess players and whether openings provide an advantage over the opponent have been researched for many years. Studies are mostly based on mathematics and analytics as a result of the analysis of grandmasters and generally the openings played in tournaments. The popularity of Network Science in chess studies has increased in recent years (Marzo & Servedio, 2023). As a result of both the visual analysis opportunity offered by complex networks and the more different and complex research opportunities, the relationship between players and openings becomes more understandable thanks to Networks.

Leveraging the dataset provided by Lichess, Marzo, and Servedio created a bipartite network between the players and openings in order to investigate the relationship between them. Researchers, who detect different opening types in the network using a non-specified community detection algorithm, assigned different complexity scores to the games using the "Economic Fitness and Complexity Algorithm" (Marzo & Servedio, 2023). As a result of their analysis, they observed that the playing styles of players with different Elo scores are clearly separated from each other. By assigning a complexity score to each opening style, the researchers found that the higher the player's Elo score, the less they tended to use low-complexity games. In other words, it is claimed that the mastery of the players and the probability of using the openings that were used less frequently show a positive correlation with each other (Marzo & Servedio, 2023).

2.2 Structure Constrained By Metadata In Networks Of Chess Players

Almeira et al. (2017) investigated the topological characteristics of the networks by conducting an analysis player's level of play as node metadata. This study, in which the chess community is handled as a data network, reveals how the playing styles of players with different experiences differ from each other. Players were treated as "nodes" and the Elo change between two opponents at the end of each game is graphed with edges. In the study, in which the community detection algorithm developed by Newman and Clauset was used, statistical analyzes were performed on values such as clustering coefficient and correlations. As a result of the analysis by Almeira et al. (2017), the players' elo values were found to be a big factor in community building. When players with higher Elo encounter players with much lower Elo points than themselves, if they beat them, the Elo points they earn will be quite low. However, if they are defeated, the Elo points they lose are considerably higher than the Elo points they lose when they play with someone of their own level. For this reason, it has been observed that chess players with high Elo scores avoid playing with low Elo players, creating an elite stratum among the chess community. Studies also show that a very small fraction of players contributes to the majority of the elo flow (Almeira et al., 2017). This study reveals how hierarchical the chess community has among itself.

3. Analysis and Findings

3.1 The Visual Analysis of Bipartite Network

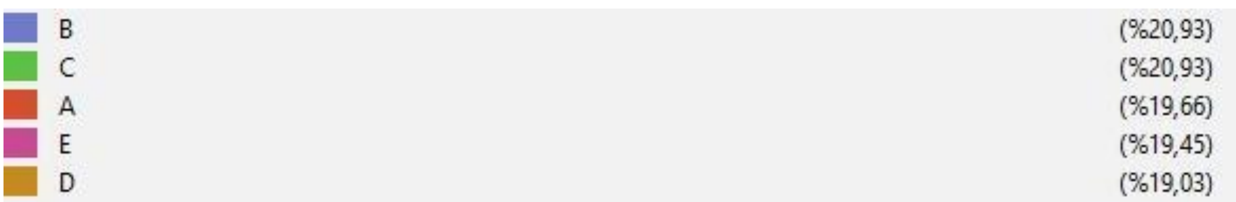


FIGURE 2

A bipartite network was created to better examine the link between players and their openings, which are the two types of nodes in the network. It is examined which opening type the players used the most according to their ELO scores. The opening types are categorized into five different colors, which can be seen in Figure 2. B openings are represented in blue, C in green, A in red, E in pink, and D in mustard color. In addition, looking at Figure 2, it can be concluded how often these gameplay styles are used among all players in the league.

During the analysis of the Bipartite Network, the network containing all the players and games in the league was visualized in three different ways. In the first, the openings used or encountered by players with Elo scores between 1201 and 1600 were resized according to how often they used them. The sizes of these opening types have been increased in proportion to the frequency of use of players in the given Elo range. In addition, for a better understanding of visual analysis, how often players in the targeted Elo range use these opening types is revealed on the histogram. This process is completed in the same way for players whose Elo range was 1601 and 2000, and 2001 and 2400.

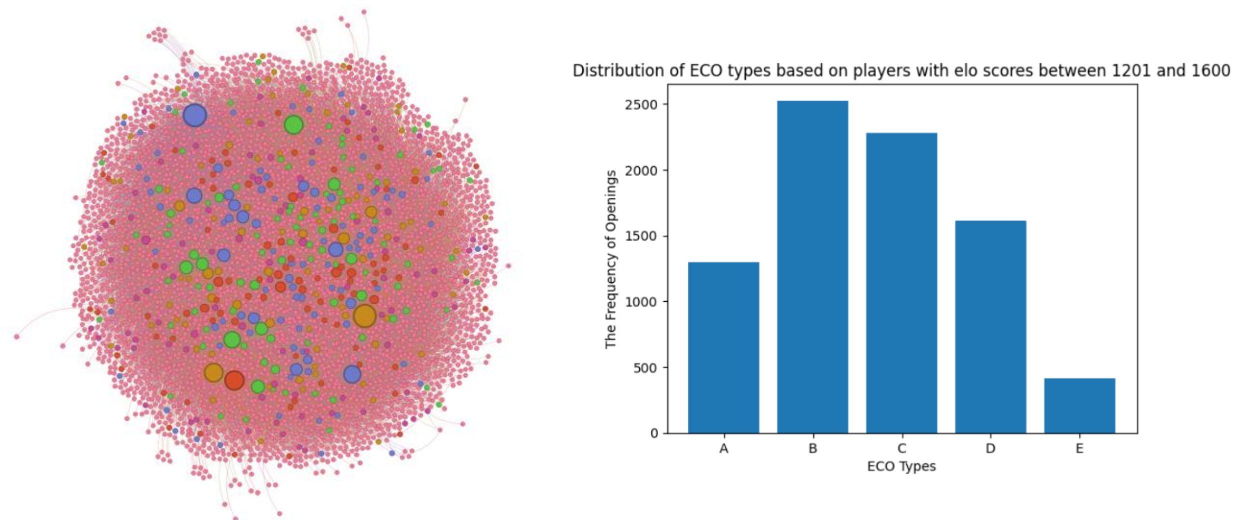


FIGURE 3

As can be seen in Figure 3, the most preferred opening types by players with an Elo range between 12001 and 1600 are B, C, D, A, and E, respectively. Except for the E-type opening, it is possible to say that some openings are more dominant than others in all other opening types. When the network is considered in a visual context, E-type openings are distributed almost equally. On the other hand, It is observed that there are dominant openings in other opening forms.

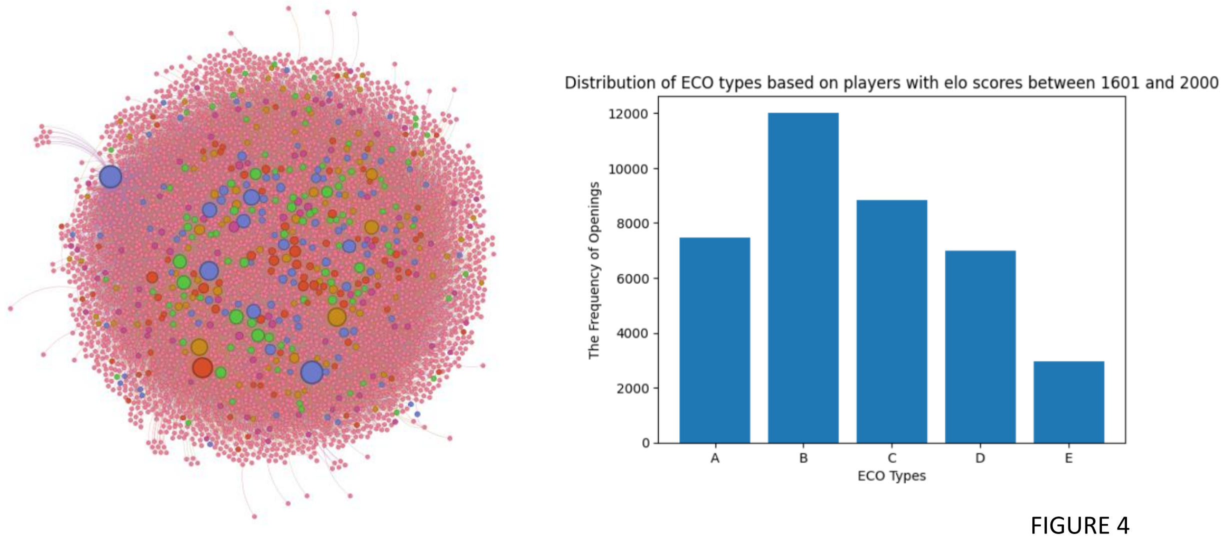


FIGURE 4

Looking at Figure 4, it is possible to say that the popularity of opening types varies between players between 1601 and 2000 Elo points, compared to players with 1201 and 1600 Elo points. Although A and C opening types are still the most used types, A-type openings have become more popular than D-type openings. Visually, it is also possible to observe dominant opening types, similar to the previous visualization.

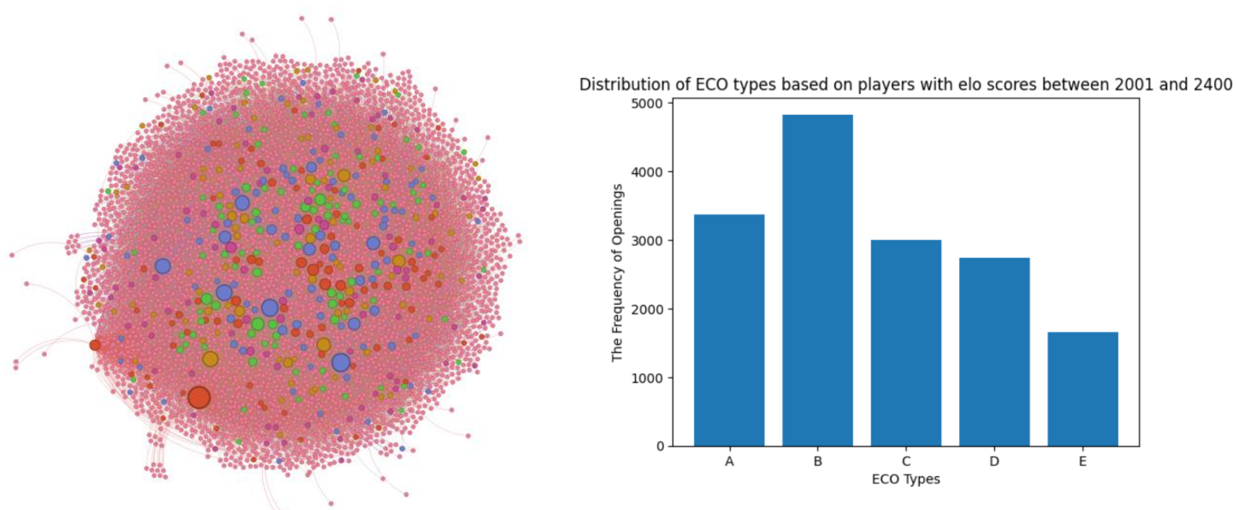


FIGURE 5

When looking at chess players with an Elo score between 2001 and 2400, the general game trend becomes more clear. Although B-type openings continue to show themselves as the most popular opening type, A-type openings keep on with their increasing trend and take their place as the second most used opening type among players with high Elo scores. Also, looking at Figures 3, 4, and 5, it would not be wrong to say that the use of E-type openings increases as Elo scores increase.

Considering the networks used at this stage, it is seen that type B openings are always the most popular opening format, regardless of the Elo score. It is possible to say that A-type openings are an opening type that increases their popularity exponentially as the Elo score increases. Such that while it is the fourth most popular opening type among players in the 1201 and 1600 Elo range, the 1601 and 200 Elo score range has become the third most used opening type by the players. Looking at the players in the 2001 and 2400 Elo score range, it is in second place. Based on this, it has been observed that the frequency of A-type openings increases in direct proportion to the Elo scores. The same positive acceleration is also observed in E-type openings. However, this acceleration is not as much as in A-type openings.

3.2 The Mathematical Analysis of Bipartite Network

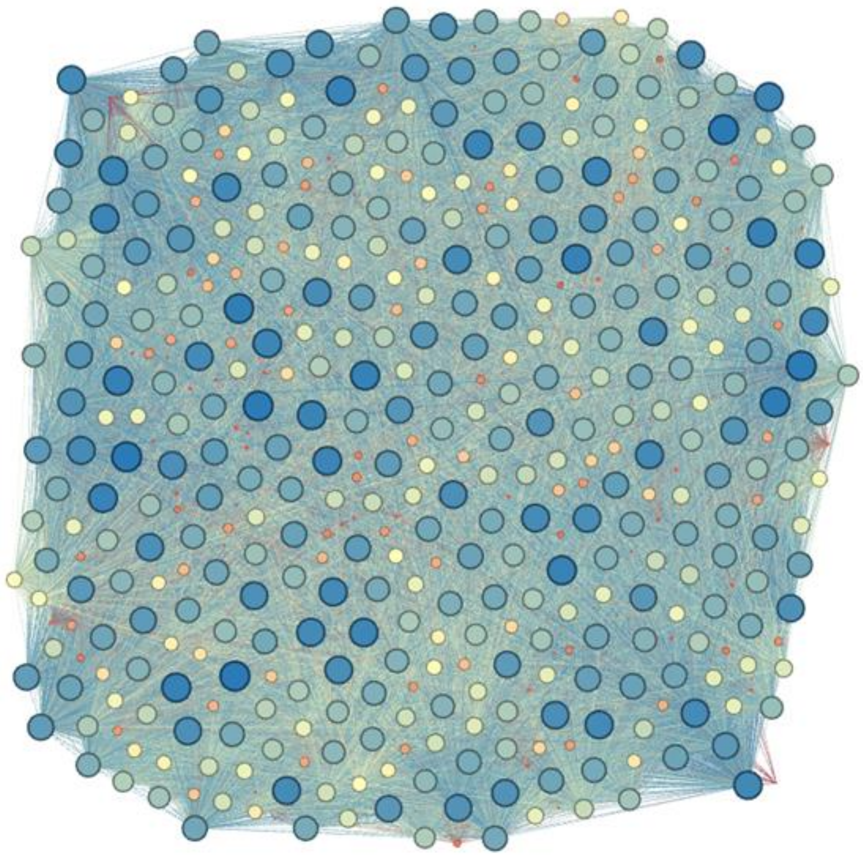


FIGURE 6 Visualization of network based on node degrees

In Figure 6, the network of openings can be seen. It is projected from the original "player-eco type" bipartite network. The colors and sizes of nodes are determined based on the degree value of each node. The nodes with higher degree are larger in size and their color is closer to blue. As the node degree decreases, the node color changes from blue to white and then to red.

In this network, there are 473 nodes and a total of 68019 edges. The general degree distribution of the network can be seen in Figure 7. The distribution is gathered mostly between 250 and 400. When we group opening types into different categories, the B-type of openings have the highest degree distribution in the network, and the E-type of openings have the lowest degree distribution in the network. Considering this distribution, it can be inferred that the player mostly preferred “B” type of openings to start the match. The distribution based on opening classes can be seen in Figure 7.

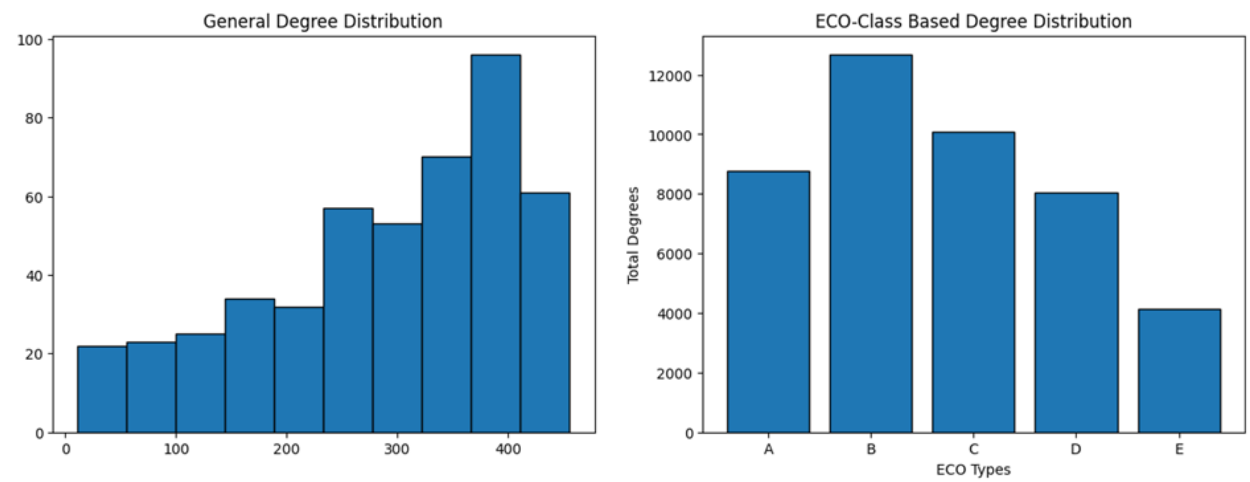


FIGURE 7 General and Class-based Degree Distribution

In Figure 8, the degree distribution of openings can be seen with their names. For the sake of simplicity, 20 nodes with the highest degree have been visualized. From the graph, it can be seen that the top 3 openings are from the B-type which has the highest degree distribution in general. More specifically, there are 3 from A-type, 7 from B-type, 5 from C-type, 5 from D-type, and 0 from E-type. As in general distribution, B-type of nodes dominate the top 20 nodes.

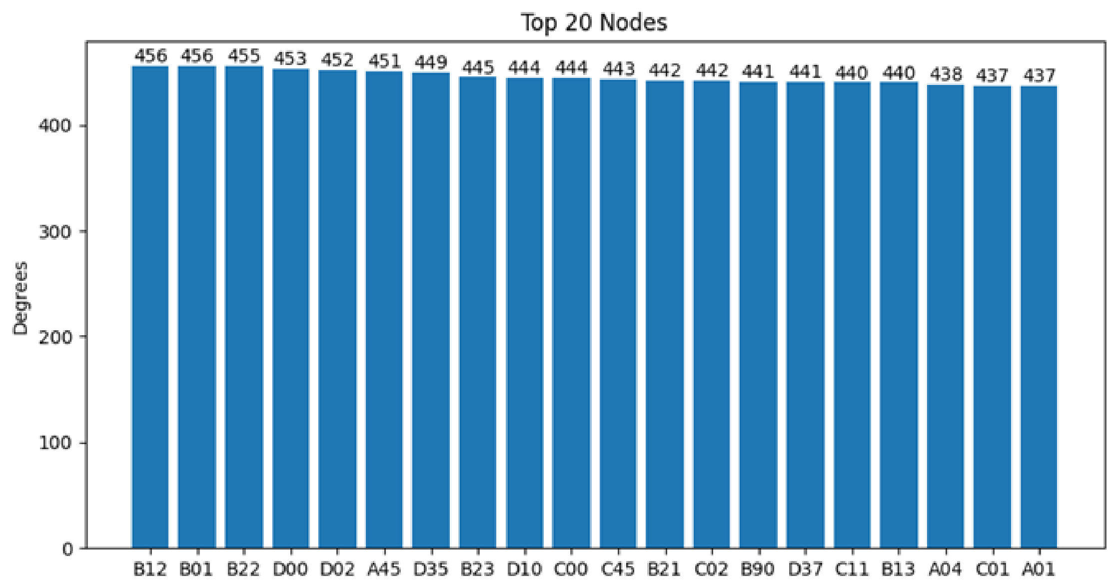


FIGURE 8 Degree Distribution Top 20 Nodes

To analyze the network in terms of centralities; closeness, degree and betweenness centralities have been used. For sake of simplicity, top 20 nodes from each centrality measurement have been chosen and visualized. In Figure 8, degree centrality measurement can be seen. Since degree centrality is based on degree distribution, the top 20 nodes are the same as the graph in Figure 8.

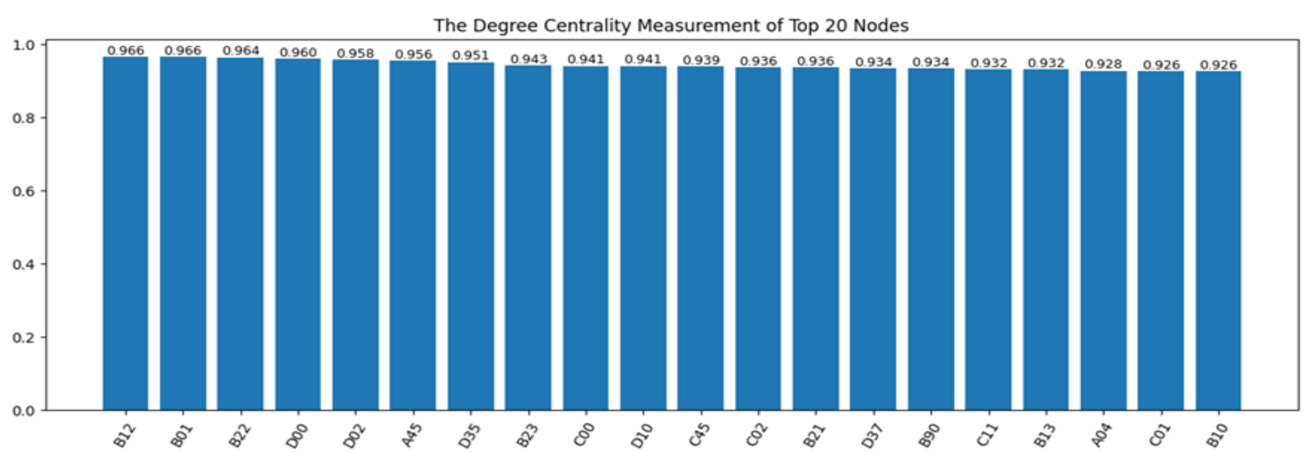


FIGURE 9 Degree Centrality Top 20 Nodes

In Figure 10, the closeness centrality measurement can be seen. When we inspect the graph, it can be seen that the names of openings in closeness measurement and the names of openings in degree centrality measurement are the same and again the top 20 nodes are dominated by the B-type of openings. This analysis has been made by using NetworkX. By using Gephi, we can confirm the measurement. The analysis result of Gephi can be seen in Figure 11. By comparing the results of Gephi and Networkx, we can say that our previous inferences are valid.

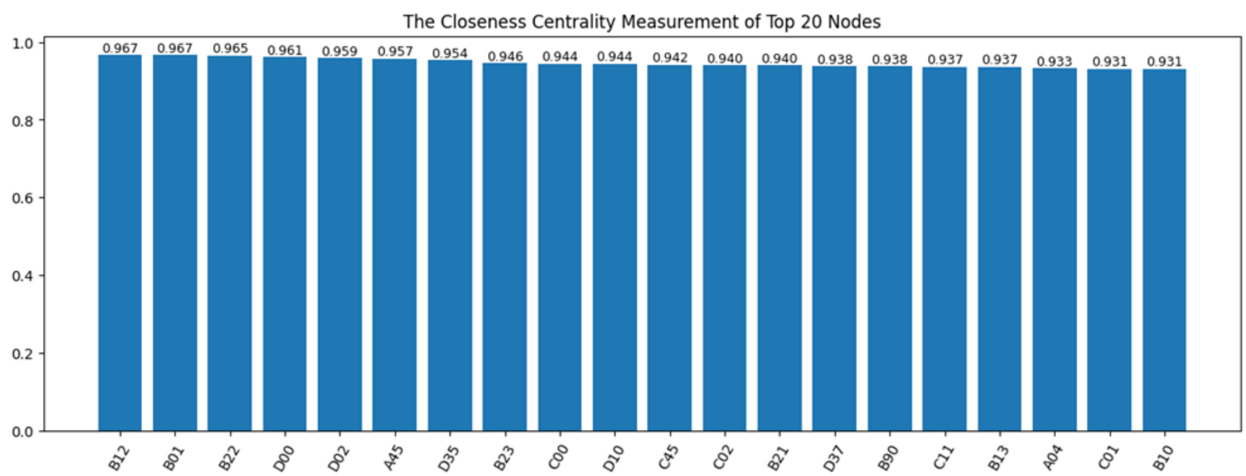


FIGURE 10 Closeness Centrality Top 20 Nodes (NetworkX)

Id	Closeness Centrality
B12	0.967213
B01	0.967213
B22	0.965235
D00	0.961303
D02	0.95935
A45	0.957404
D35	0.953535
B23	0.945892
D10	0.944
C00	0.944
C45	0.942116
C02	0.940239
B21	0.940239
D37	0.93837
B90	0.93837
C11	0.936508
B13	0.936508
A04	0.932806
C01	0.930966
B10	0.930966
A01	0.930966

FIGURE 11 Closeness Centrality Top 20 Nodes (Gephi)

In Figure 12, the betweenness centrality measurement can be seen. When we inspect the graph, it can be seen that there are two A-type, seven B-type, five C-type and six D-type of openings. In a nutshell, betweenness measurement is dominated by B-type of openings like other centralities.

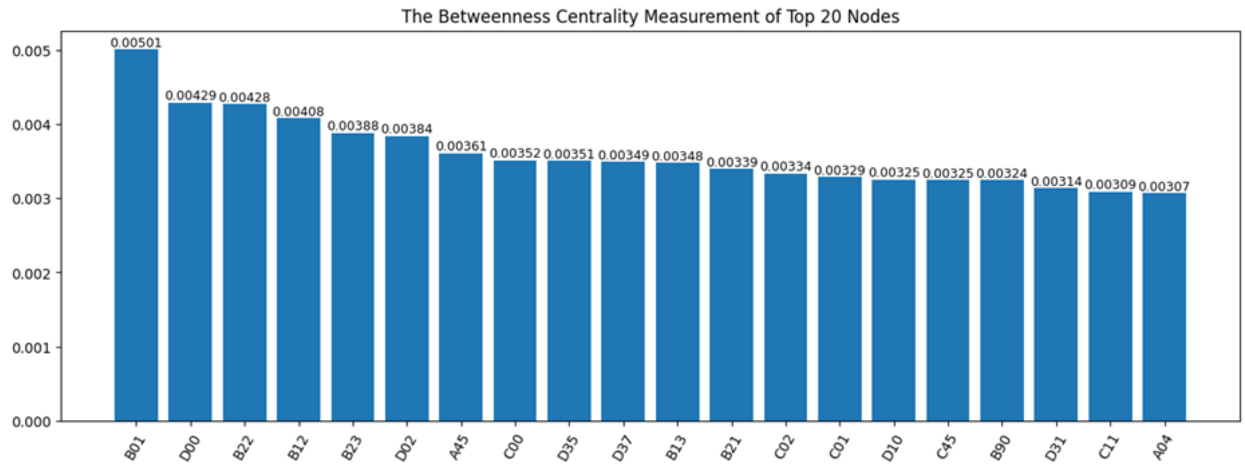


FIGURE 12 Betweenness Centrality Top 20 Nodes

Other than centrality measurements, some distance related measurement has been done. The result and type of these measurement can be listed as follows:

Radius Of Network	2
Diameter Of Network	3
Periphery Of Network	‘A93’ – ‘E59’ – ‘D57’
Centre Of Network	‘B12’ – ‘B01’ – ‘B22’
Average Path Length	1.39

FIGURE 13

Finally, clustering coefficients are performed. The density of the network is calculated, clustering coefficients of nodes and the average clustering of the whole network. The density value of the network is 0.609. It can be inferred that the network is closer to being a dense network. In Figure 12, the clustering coefficient distribution can be seen. For the sake of simplicity, the top 20 nodes have been visualized. When we inspect the graph, there are five A-type, four B-type, three C-type, six D-type and two E-type. Unlike centrality measurements, the clustering measurement is dominated by D-type of nodes and as expected E-type of nodes has the lowest count. By using nodes' clustering coefficients, it has been calculated the average clustering of the whole network. The result is 0.845. It can be said that the network is highly clustered.

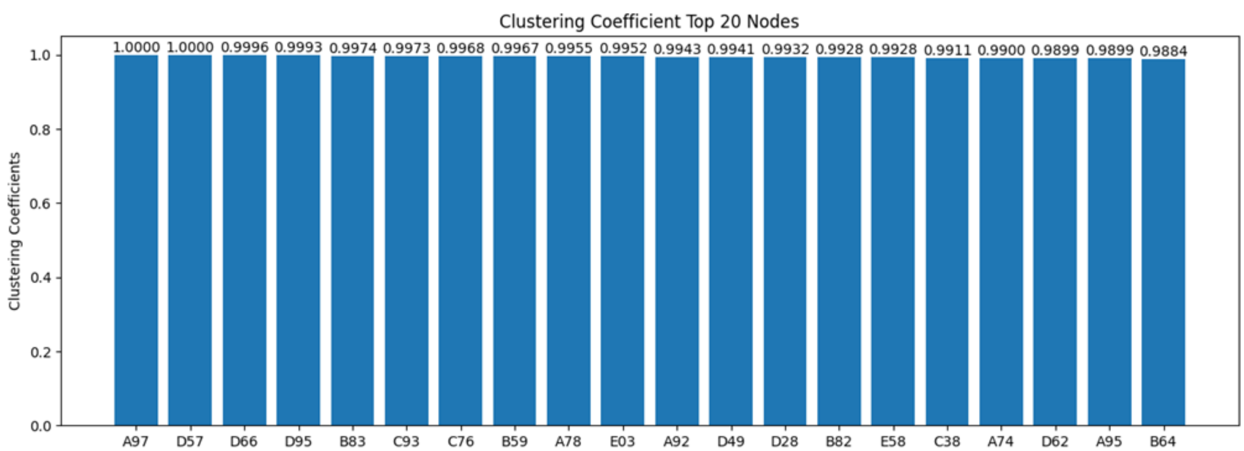


FIGURE 14 Clustering Coefficient Top 20

3.3 Community Detection

A community detection algorithm is performed in order to emphasize the natural similarities between the openings instead of the standard ECO separation. ECO types are commented and designed by humans which leads to a natural vulnerability of bias. In community detection, the opening projections of the bipartite network are employed. With the help of the Louvain method, four different communities are gathered, which can be seen in the figures below. On top of the community clustering, we divide the data into four different groups of ratings: 800-1200, 1201-1600, 1601-2000, and 2001-2400.

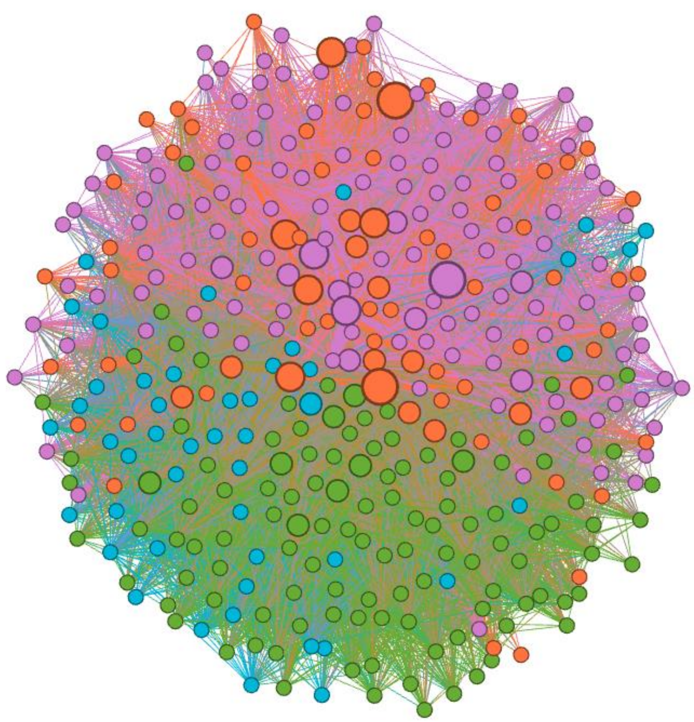


FIGURE 15: 800-1200 rating group, colorized by the community detection

In Figure 15, it can be seen that lower-rated players tend to play a variety of openings and although there are some popular ones, they also play very unpopular and dubious openings. There is not much of a centralization around the popular openings or communities.

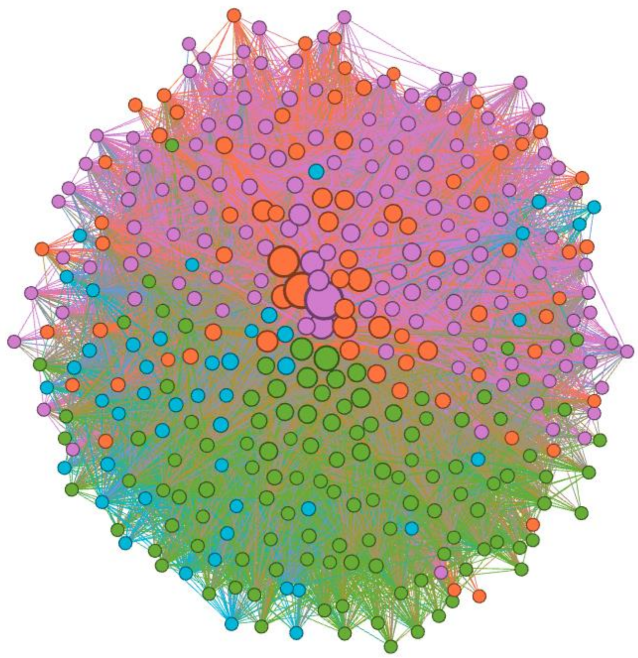
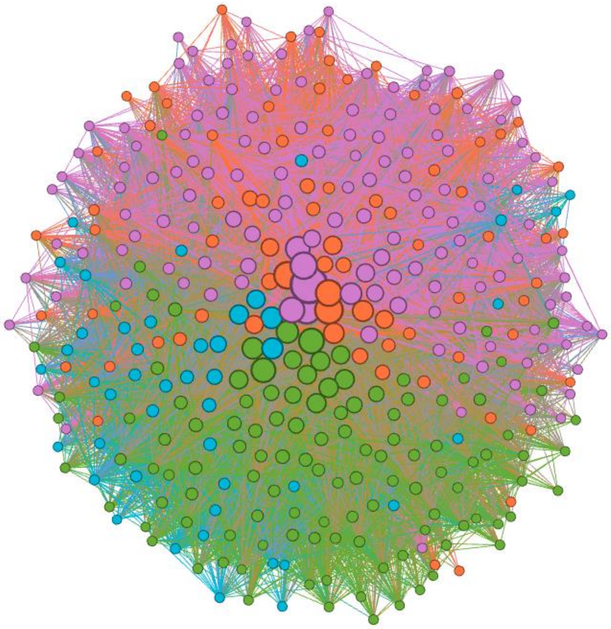


FIGURE 16: 1201-1600 rating group, colorized by the community detection

In Figure 16, there is a visible centralization by the popularity of the openings. Despite the fact that there is still a variance, players seem to start to play more sound openings which are also suggested by the professionals.



In Figure 17, variety drops even more and the most popular openings become visible. Players choose some openings and play only them. They tend not to prefer preparing a big opening repertoire.

FIGURE 17: 1601-2000 rating group, colorized by the community detection

In Figure 18, a strange phenomenon is observed. Although there has been a steady increase in the centralized openings since the ratings of 800, much higher the rating is, the more variety the openings have. Especially

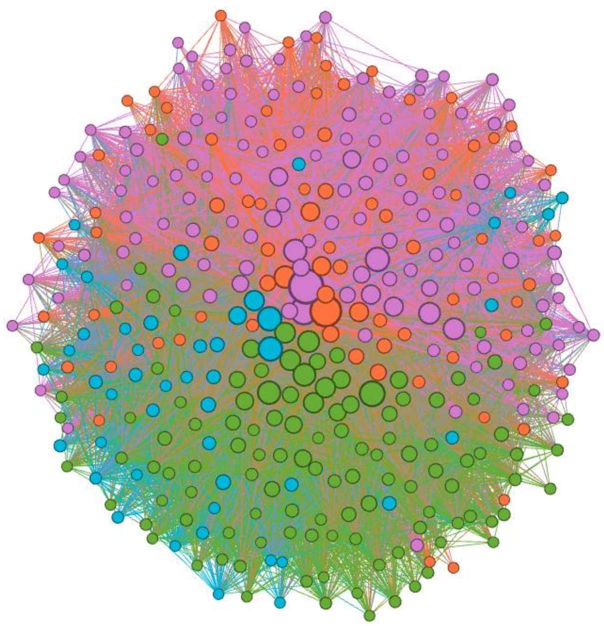


FIGURE 18: 2001-2400 rating group, colorized by the community detection

in the last segment, 2001-2400, players tend to get rid of their habits and start to play different and unusual openings. This situation might be a significant factor that helped these players to get these high ratings.

4. Conclusion

In this research paper, a network analysis approach is employed to investigate the relationship between opening choices and player strengths in online chess. By analyzing data from the Lichess4545 online chess league, a player-opening bipartite network is constructed, and community detection methods are employed to compare players of different skill levels. The findings of the study have provided valuable insights into the strategic preferences and advantages associated with various skill levels in chess.

The analysis revealed intriguing and non-linear patterns between player strengths and opening choices. It is observed that players of higher skill levels tended to exhibit distinct opening preferences, suggesting the presence of strategic choices that align with their expertise. This finding indicates that players at different skill levels employ different strategic approaches during the opening phase of a chess game. Furthermore, the research demonstrated that network analysis is an effective tool for capturing the interplay between players and their preferred opening moves, providing a comprehensive understanding of the complex dynamics of online chess gameplay.

The insights gained from this study have practical implications for both chess players and game analysts. Understanding the strategic preferences of players at different skill levels can guide training strategies, allowing players to focus on openings that align with their strengths and improve their overall gameplay. Additionally, game analysts and chess coaches can leverage these findings to develop tailored training programs that address the specific needs of players at different skill levels, enhancing their performance and strategic decision-making abilities.

While the study sheds light on the relationship between opening choices and player strengths, there are several avenues for future research. Expanding the analysis to include a larger dataset from multiple online chess platforms and incorporating additional variables, such as time controls and game outcomes, would provide a more comprehensive understanding of the dynamics between openings and player strengths. Furthermore, exploring the impact of specific opening choices on subsequent game development and outcomes could offer valuable insights into the strategic advantages and disadvantages associated with different openings.

In conclusion, the research contributes to the chess literature by providing empirical evidence of the association between opening choices and player strengths in online chess. By employing network analysis techniques, we have deepened our understanding of the complex interplay between players and their preferred opening moves. The findings from this study have practical implications for chess players and game analysts, and they highlight the potential of network analysis in studying strategic games and complex systems. Ultimately, by gaining insights into the strategic choices made during the opening phase, players can enhance their gameplay and increase their chances of success in online chess competitions.

References

About lichess.org. (n.d.). Retrieved from lichess.org: https://lichess.org/about

Almeira, N., Schaigorodsky, A. L., Perotti, J. I., & Billoni, O. V. (2017). *Structure Constrained By Metadata In Networks Of Chess Players.* Nature.

Marzo, G. D., & Servedio, V. D. (2023). *Quantifying the complexity and similarity of chess openings using online chess community data.* Nature.