Chess Through Space, Time, and Data: A Statistical Exploration

Idan Yankelev, Michael Fire
Department of Software and Information Systems Engineering
Ben-Gurion University of the Negev, Be'er-Sheva, Israel

***Abstract*—Chess has undergone significant transformations over time, influenced by evolving player strategies, regional dominance, and statistical trends in gameplay. This study presents a data-driven analysis of Chess using historical game records, focusing on three key dimensions: Space, examining the geographical distribution of chess activity; Time, tracing trends in openings, player rankings, online vs. offline play, and statistics, exploring correlations between openings, Elo ratings, and game outcomes. Using data from The Week in Chess (TWIC) archive, we explore and analyze these dimensions and present various visualization tools to uncover meaningful insights. Our findings highlight shifts in opening popularity, the rise of dominant players, regional strengths in Chess, and statistical relationships between game elements, providing a comprehensive view of how Chess has evolved globally and offering valuable insights for players, analysts, and enthusiasts alike. Our code is provided in https://github.com/IdanYankelev/Chess-Analysis**

***Index Terms*—Chess Data Analysis, Chess Openings Trends, Elo Correlation, Temporal Chess Patterns, Geographical Chess Distribution, Statistical Chess Insights, Chess Strategy Evolution**

I. Introduction

Chess, with its rich history spanning over a millennium [1], has continually evolved, reflecting shifts in strategy, player dynamics, and global participation. The advent of comprehensive game databases and advanced analytical tools has opened new avenues for understanding these evolutionary patterns.

One interesting area of study is the analysis of chess openings [2]. These initial moves set the stage for the game and have been the subject of extensive research [3]–[6] which have shown a correlation between a player's familiarity with specific openings and their overall performance. For instance, an analysis of 20,000 games from the Lichess platform [1] revealed that higher-rated players tend to adhere more closely to established opening moves, suggesting a strong link between opening knowledge and player success [6].

Another intriguing aspect is the geographical distribution of chess talent. Historically dominated by specific regions (e.g., Russia), the landscape of elite Chess has diversified over time, with the current top players being from Norway, the USA, and India (according to Chess.com [2]). According to [7], countries like India have recently seen a surge in chess popularity and talent development, contributing to a more global representation among top players.

Lastly, the Elo rating system remains a cornerstone in assessing a chess player's strength. Developed by Arpad Elo [8], this system quantitatively measures a player's skill based on game outcomes against other rated players. Understanding the distribution and progression of Elo ratings offers insights into competitive dynamics and the effectiveness of training methodologies across different regions.

In this study, we conduct a comprehensive analysis of chess games, focusing on three primary dimensions:

1) **Temporal Trends**: Examining how opening preferences, player performance, and game outcomes have evolved, particularly with the rise of online platforms.
2) **Geographical Patterns**: Analyzing the distribution of chess talent globally, exploring factors contributing to regional strengths and the emergence of new chess hubs.
3) **Correlation Analysis**: Investigating chess openings' prevalence and success rates across different player ratings and over time.

Utilizing data from *This Week in Chess* (TWIC) archive [3], we employ advanced statistical methods and visualization techniques to uncover patterns and trends. This research aims to provide a nuanced understanding of the factors influencing chess evolution, offering valuable insights for players, coaches, and enthusiasts seeking to navigate the ever-changing landscape of competitive Chess.

The remainder of this paper is structured as follows:

- Section II reviews related work.
- Section III describes our dataset and methodology.
- Section IV presents our findings
- Section V discusses key insights from our findings
- Section VI concludes our work and discusses future directions.

II. Related Work

The relationship between a player's knowledge of chess openings and their performance has been extensively studied. Comer [6] analyzed 20,000 games from the Lichess platform, revealing that higher-rated players tend to follow established opening moves more closely, indicating a positive correlation between opening knowledge and player success.

Olson [9] conducted a comprehensive analysis of over 650,000 chess tournament games to explore the evolution of opening popularity. The study found that while the King's

[1]https://lichess.org/

[2]https://www.chess.com/players

[3]https://theweekinchess.com/twic

Pawn Opening dominated in the 1850s, opening strategies were diversified over time, with over 1,000 unique openings recorded by 2014. This shift suggests an increasing complexity and depth in opening preparations among players.

Blasius and Tonjes [10] investigated the frequency distribution of opening moves and discovered that it follows a power-law distribution, adhering to Zipf's law. This finding implies that a few openings are extremely popular, while a long tail of less common openings exists, reflecting the vast diversity in players' strategic choices.

Chowdhary et al. [11] analyzed over 120 million games, finding that skilled players often specialize in specific openings, while beginners tend to experiment more. This specialization reflects a deeper understanding and preparation in particular lines, contributing to consistent performance.

III. Methods

This section outlines the methods for analyzing chess trends across time, space, and statistical dimensions.

A. Datasets

This study's primary dataset is sourced from the *The Week in Chess* (TWIC) archive, which provides a comprehensive repository of historical and contemporary chess games. Each record in the dataset contains essential metadata, including:

- **Player Information:** Names, Elo ratings, and nationality.
- **Game Details:** Move sequences, results, and timestamps.
- **Opening Classification:** Encoding using the Encyclopedia of Chess Openings (ECO) codes [4].
- **Geographical Context:** Tournament locations.

To enrich the analysis, additional data sources, such as player rankings and nationalities from Chess.com [5] and geographical data from folium [6], are incorporated.

B. Data Preprocessing

Prior to the analysis, the dataset underwent a preprocessing pipeline to ensure consistency and integrity:

- **Data Cleaning:** Removal of duplicate records, faulty entries (e.g., incorrect timestamps), and incomplete data.
- **Standardization:** Normalization of player names and aggregation of similar opening variations under broader categories (e.g., variations of the Queen's Gambit).
- **Filtering:** Exclusion of non-rated players.

C. Analytical Framework

1) Temporal Trends Analysis: To capture the evolution of chess strategies, we analyze game data chronologically. Games are grouped by the event year, and we track the frequency of different openings over time. This allows us to observe shifts in opening preferences and play styles (defensive vs. offensive) variations across different eras. We also compare offline and online play trends to examine how digital platforms have influenced chess strategies.

[4]https://www.365chess.com/eco.php

[5]https://www.chess.com/

[6]https://github.com/python-visualization/folium

2) Geographical Distribution: To explore the global chess landscape, we analyze the distribution of grand master players and game occurrences across different world regions. By aggregating player Elo ratings by country, we identify chess powerhouse countries (e.g., Russia) and the emergence of new chess hubs (e.g., India) alike.

3) Statistical Correlation Analysis: We conduct correlation studies to examine relationships between game attributes. Specifically, we explore:

- The impact of opening choices and Elo differences on game outcomes, stratified by Elo rating brackets (1400-1600, 1600-1800, 1800-2000, 2000-2200, 2200-).
- The evolution of specific opening effectiveness over different periods.
- The connection between the match's location and game outcome.
- The role of color advantage (White vs. Black) across different rating distributions.

IV. Result Analysis

This section presents our findings based on our statistical analysis of chess games across temporal, geographical, and statistical dimensions. Our results reveal notable trends in chess openings, global player distribution, performance variations, and game outcomes.

A. Temproal Trends

Figure 1 presents the ranking trends of the top 10 chess players based on their ELO ratings over several years (2012-2019), with a unique color representing each player.

As can be seen and is a well-known fact among chess enthusiasts, Magnus Carlsen has been considered the top player since becoming World Champion in 2013. Carlsen has maintained the highest ELO rating and consistently held the top ranking. Meanwhile, the rankings of other top players fluctuate significantly over time, reflecting the intense competition at the elite level. Players like Fabiano Caruana, Hikaru Nakamura, Levon Aronian, and others have had periods of strong performances, occasionally challenging for the top spots but never maintaining Carlsen's consistency.

Figure 2 displays the ranking trends of the top 10 chess openings over time (2010-2022), illustrating the evolving landscape of opening preferences at the highest level. Unlike the player ranking trends in Figure 1, opening choices exhibit a more structured yet dynamic evolution, with certain defensive (e.g., Sicilian defense) and offensive systems (e.g., Caro-Kann) maintaining long-term popularity while others experience significant shifts.

Classical openings like the Ruy Lopez opening, which dominated historical play, maintain a strong presence but face increasing competition from alternative setups. The French Defense and Slav Defense, traditionally reliable choices, also exhibit steadiness but remain outside the top-tier trends compared to more versatile modern defenses.

Lastly, we also wanted to explore the ratio of offline vs online games over time, illustrated by Figure 3. As can be

seen, offline games dominated from 2012 to 2019, with a steady annual increase in their number. While present, online games remained marginal in comparison. However, a drastic shift occurred in 2020, when online games saw an exponential surge, overtaking offline games for the first time and forever reshaping the chess landscape.

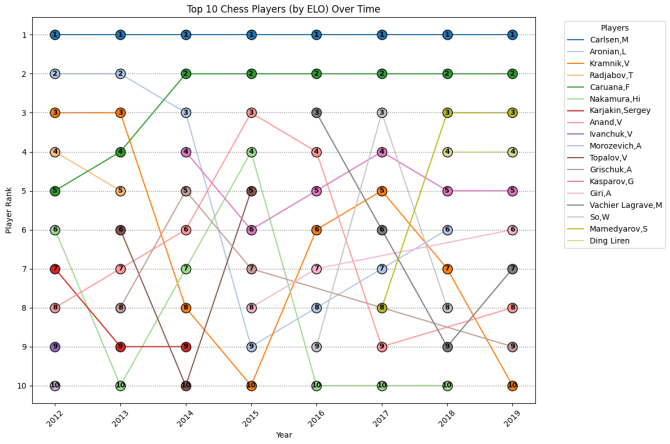

Fig. 1. Temporal trend of top 10 players (by Elo) over the 2012-2019 era.

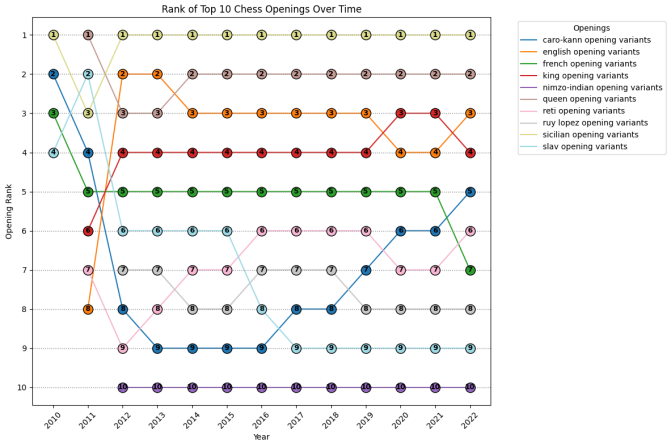

Fig. 2. Temporal trend of top 10 opening categories over the 2010-2022 era.

B. Geographical Distribution

Figures 4 and 5 provide a geographical perspective on chess activity worldwide.

Figure 4 shows the geographical distribution of chess matches, where countries are colored based on the number of games recorded, with darker colors representing more significant amounts of games. The highest densities of matches are observed in regions with established chess cultures and strong online participation, such as Europe, North America, and parts of Asia. Many African and smaller nations are either sparsely represented or absent, highlighting disparities in chess accessibility and engagement.

In contrast, Figure 5 focuses on the number of grandmaster-level players per country, where countries are colored based

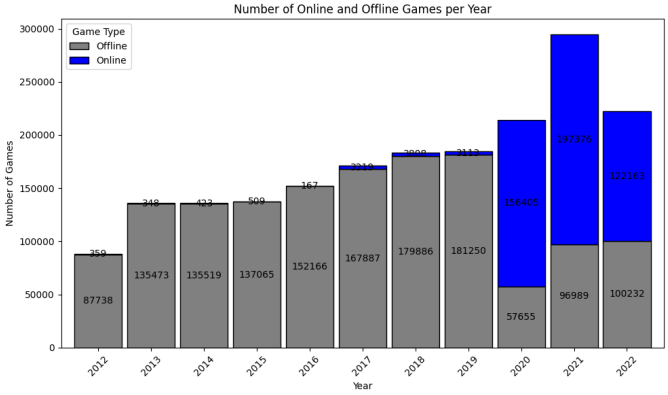

Fig. 3. Temporal trend of offline vs. online games over the 2012-2022 era.

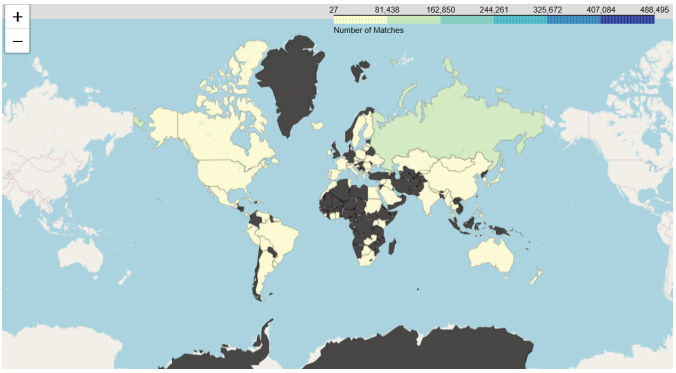

Fig. 4. Geographical distribution of grandmaster chess events worldwide.

on the number of local grandmasters, with darker colors representing more significant amounts of grandmasters. Russia and other traditional chess powerhouses in Eastern Europe display the highest concentrations of elite players, reinforcing their historical dominance.

Both maps underscore a key trend: while online play has expanded global chess engagement, elite talent production remains concentrated in historically strong chess regions.

C. Statistical Correlations

Figure 6 examines the relationship between openings and game outcomes, distinguishing between victories for White, victories for Black, and draws. The color gradient represents the degree of Correlation, with red shades indicating a positive correlation and blue shades indicating a negative correlation. As can be seen, the Sicilian Defense and Ruy Lopez openings show very slight positive correlations with decisive games, particularly favoring White. In contrast, openings like the Scandinavian and Modern Defense openings exhibit minor tendencies toward balanced or defensive play.

Similarly, Figure 7 examines the relationship between openings and various ELO ranges, from novice to master-level players. While most openings exhibit near-zero Correlation with ELO categories, subtle trends emerge. For instance, some openings, such as the Bird's Opening, Nimzo-Indian, and Scandinavian Defense, show slight positive correlations

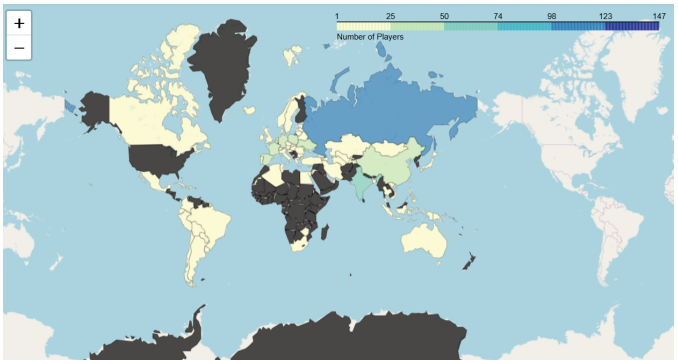

Fig. 5. Geographical distribution of grandmaster players worldwide.

with higher-rated players. In contrast, other openings, like the Catalan and Slav Defense, tend to be more common at lower levels. These correlations, although small, suggest that specific openings may be more naturally suited to players of specific skill levels, potentially due to their complexity or strategic depth.

In addition, when we calculate the Correlation between Elo Differences between white and black players and the game outcome, we observe a positive correlation of 0.48, suggesting that these differences mostly favor white victory (as expected).

V. Discussion

A. Opening Trends and Strategic Evolution

A key observation from our temporal analysis is the stability of certain defensive openings, particularly the Sicilian Defense, which consistently holds top ranking. The Caro-Kann and Nimzo-Indian Defenses openings also exhibit long-term resilience, reflecting their structural solidity and enduring theoretical relevance. They remain favored by players who prioritize positional soundness and counterattacking potential.

In contrast, offensive and hyper-modern openings show more significant fluctuations in ranking. The English opening and Reti System, which emphasize fluidity and indirect central control, experience periodic surges in popularity but lack long-term stability. Similarly, aggressive opening choices such as the King's Gambit opening exhibit relatively lower consistency at elite levels, reinforcing the trend that modern high-level play favors controlled, strategic play over early tactical imbalances.

B. Impact of Online Play

Our findings highlight a drastic shift in the landscape of chess gameplay, particularly post-2020. Prior to this period, offline games consistently outnumbered online play. However, with the onset of the COVID-19 pandemic, online games surged exponentially, surpassing offline records.

In addition to COVID-19, another key event in 2020 contributed significantly to the rise in chess interest: the release of Netflix's The Queen's Gambit show, which follows the journey of a child chess prodigy. The show sparked a cultural phenomenon, leading to increased engagement with Chess across all levels. Online platforms reported record-breaking

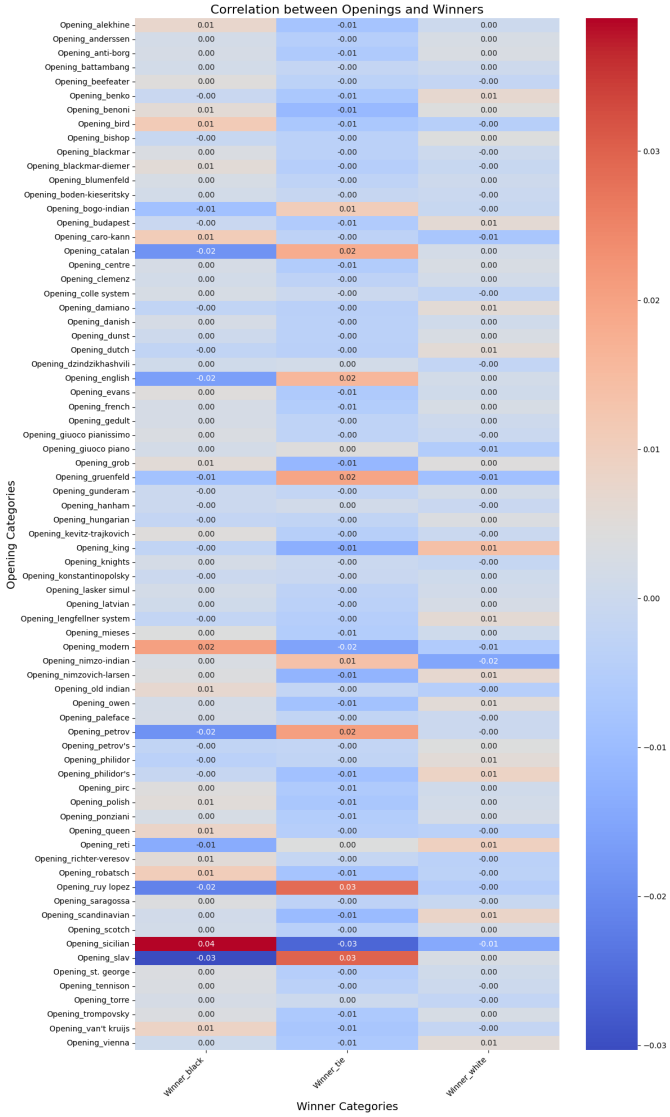

Fig. 6. Correlation between opening strategy and the match outcome.

numbers of new users, tournament participation increased, and chess-related Google searches spiked [7].

C. Geographical Distribution of Chess Strength

Our analysis shows a strong concentration of grandmaster-level players in historically dominant chess nations like Russia. The Russian chess school's influence remains evident, with a high density of grandmasters per capita. India's rapid ascent in Chess, supported by a strong digital chess presence, highlights the emergence of new powerhouses in the sport.

Interestingly, the geographical disparity between the total number of chess games played and the number of titled players suggests that while digital Chess has globalized the game, elite chess talent remains unevenly distributed. Many regions with high game volumes, such as North America, still lag in developing a proportionate number of titled players. This

[7]https://shorturl.at/gja3J

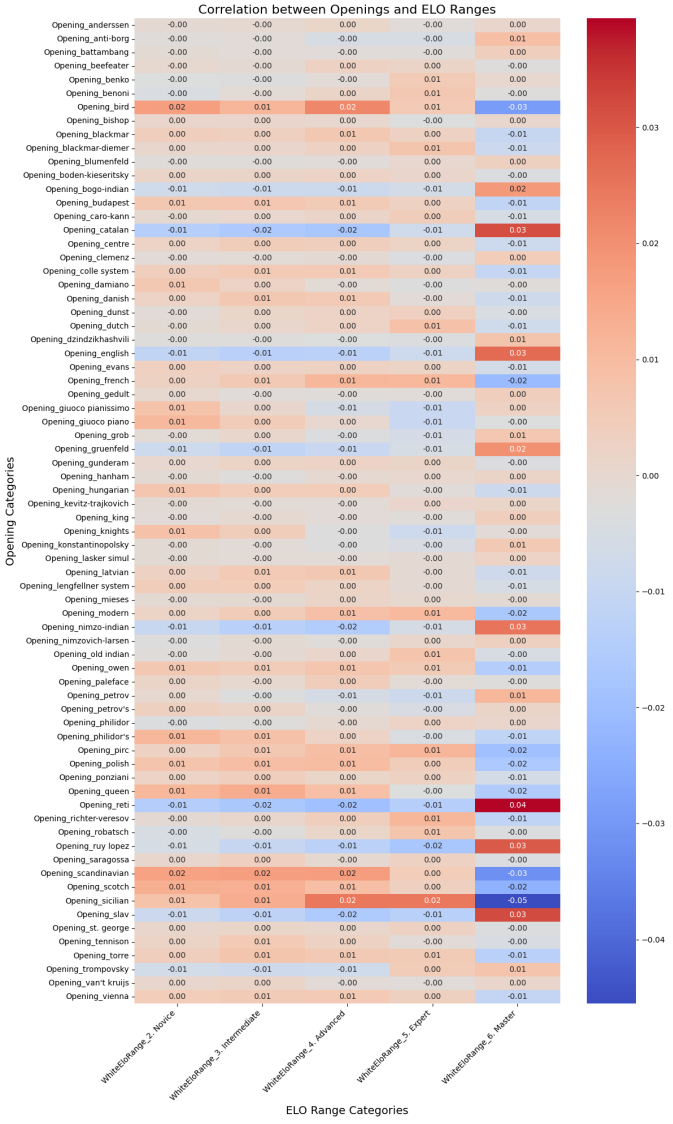

Fig. 7. Correlation between opening strategy and the player's Elo rating.

indicates that infrastructure, coaching, and access to competitive opportunities play crucial roles in producing world-class talent.

D. Statistical Correlations and Performance Factors

Our analysis reveals subtle trends regarding openings, player strength, and game outcomes. The Sicilian Defense and Ruy Lopez openings show slight positive correlations with decisive results, suggesting that these openings lead to more dynamic and non-drawing positions. On the other hand, more solid choices, such as the Caro-Kann and Slav Defense, exhibit a slight tendency toward balanced or defensive play.

The analysis of openings across different ELO ranges suggests that higher-rated players more commonly play specific openings, particularly those requiring deep positional understanding, such as the Nimzo-Indian and Scandinavian Defense openings. Meanwhile, more straightforward or aggressive openings, such as the Bird's Opening, appear more frequently in lower-rated games, potentially due to their tactical nature. These findings reinforce that while opening choice plays a role in game outcomes, the most substantial determining factor remains the midgame and endgame plays, as evidenced by Magnus Carlsen's dominance.

VI. Conclusion and Future Work

This study provided a comprehensive statistical analysis of Chess across multiple dimensions, examining temporal trends, geographical patterns, and statistical correlations in chess strategy. Our key findings highlight:

- The enduring dominance of defensive and strategic openings, while aggressive openings exhibit greater volatility.
- The impact of online Chess, with game volume shifting dramatically toward digital platforms post-2020.
- The regional disparities in chess activity, where countries with a high volume of online play do not necessarily produce a proportional number of titled players.
- The limited impact of openings on overall game outcomes reinforces the importance of midgame and endgame mastery over memorized preparation.

While this study offers valuable insights, several avenues for future research remain, such as leveraging machine learning techniques to predict a game's outcome based on different openings and game positions and investigating how different time formats (blitz, rapid, classical) influence the examined trends and overall player success.

Acknowledgment

We want to express our gratitude to ChatGPT for its assistance in proofreading and rephrasing suggestions to write our report in a more formal tone.

References

[1] H. A. Davidson, *A short history of chess*. Crown, 2012.
[2] B. Pandolfini, *Chess Openings: Traps and Zaps*. Simon and Schuster, 2013.
[3] K. Raghav and L. Ahuja, "Chess opening analysis using dbscan clustering and predictive modeling," in *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2024, pp. 1–5.
[4] M. Levene and J. Bar-Ilan, "Comparing typical opening move choices made by humans and chess engines," *The Computer Journal*, vol. 50, no. 5, pp. 567–573, 2007.
[5] R. Arabacı, "An investigation into the openings used by top 100 chess players," *Sport Sciences*, vol. 2, no. 2, pp. 1–11, 2007.
[6] A. Comer, "Relationship between chess opening knowledge and player rating," *adambcomer.com*, 2023. [Online]. Available: https://adambcomer.com/blog/chess-analysis/
[7] J.-L. Cazaux and R. Knowlton, *A world of chess: its development and variations through centuries and civilizations*. McFarland, 2017.
[8] A. E. Elo and S. Sloan, "The rating of chessplayers: Past and present," 1978.
[9] R. Olson, "A data-driven exploration of the evolution of chess: Popularity of openings over time," 2014. [Online]. Available: https://randalolson.com/2014/05/26/a-data-driven-exploration-of-the-evolution-of-chess-popularity-of-openings/
[10] B. Blasius and R. Tönjes, "Zipf's law in the popularity distribution of chess openings," *arXiv preprint arXiv:0704.2711*, 2007. [Online]. Available: https://arxiv.org/abs/0704.2711
[11] S. Chowdhary, I. Iacopini, and F. Battiston, "Quantifying human performance in chess," *arXiv preprint arXiv:2207.07780*, 2022. [Online]. Available: https://arxiv.org/abs/2207.07780