FIDE Chess Network

Kristijan Breznik\textsuperscript{1} and Vladimir Batagelj\textsuperscript{2}

\textsuperscript{1} International School for Social and Business Studies, Celje, Slovenia
\textsuperscript{2} Faculty of Mathematics and Physics, Ljubljana, Slovenia

Abstract: The game of chess has a long and rich history. In the last century it has been closely connected to the world chess federation FIDE\textsuperscript{1}. This paper analyses all of the rated chess games published on the FIDE website from January 2008 to September 2010 with network analytic methods. We discuss some players and game properties, above all the portion of games played between the world’s best chess players and the frequency of games among countries. Additionally, our analysis confirms the already known advantage of white pieces in the game of chess. We conclude that some modifications in the current system should appear in the near future.


Keywords: Chess, FIDE, Network Analysis, Elo Rating System.

1 Introduction

Chess is arguably one of the oldest and most popular mind games and is commonly believed to have many predecessors. Its inception had already occurred in the sixth century with a game called chatrang played in India and Persia (see Murray, 1913). By the year 1000 the game was known throughout Europe, where some major changes were made, especially in the 15th century in Italy and Spain. Today’s exact rules were first used in the early 19th century and the first modern chess tournament was held in London in 1851. The first official world chess champion, German master Johannes Zukertort, was crowned in 1894. For approximately twenty years German-speaking players were dominant, then Jose Raul Capablanca, a Cuban prodigy, won the chess crown. After his dominance a period of more or less great Soviet superiority began with only two non-Soviet champions; a Dutch player Max Euwe and Robert James Fischer from the USA. After 30 years, the current world chess champion Visnawathan Anand comes from India, from a non-Soviet country as well.

\textsuperscript{1}Fédition Internationale des Échecs
In the last one hundred years chess has been closely connected to the world chess federation FIDE, founded in 1924 in Paris and recognised by the International Olympic Committee as an International Sports Federation in 1999 (see FIDE homepage, 2010). FIDE is responsible for the organisation of many important tournaments, among other competitions for the World Chess Championship title and the Chess Olympiad. It is also its responsibility to calculate the rating points — a measure of the strength of each chess player. Chess has made a very important contribution to strength measuring of players in various sports. The mathematician Arpad Imre Elo (1903-1992) is considered as the founder of the chess rating system and even after over 40 years this system is almost unchanged and still named after him. However, recent data are more abundant. More powerful computers are available and it seems there will be some improvements in this legendary rating system. The most needed improvements include consideration of the colour of pieces in determining the expected result of each game. We will focus on this issue as well.

In this paper we will analyse all of the FIDE rated games played from January 2008 to September 2010 with network analytic methods. A more detailed description of players and game properties will follow in the next section. We believe that the network analysis is a suitable approach for the extraction of a group of players (and/or countries) who play against each other more often, and it is also an appropriate tool for a further detailed analysis of their game incidence. We are going to show that the best chess players consequently play against each other far more often than against other lower rated opponents. There are various reasons why the best players in the world feel uncomfortable playing against objectively weaker opponents. One of the recent main reasons are computers which have taken over the supremacy in the game of chess; not only as an invincible opponent but also as a faithful servant for the game’s preparation which is often more important than pure chess skills. The following obstacle for the best players to play against less successful opponents is potentially a high loss of rating points which are crucial for chess professionals. Namely, chess clubs choose and consequently pay for their professional players according to their current placement on the rating list.

The game of chess is popular in all continents and another view we are interested in is the frequency of games played among countries. We are going to try to answer various questions, e.g. in which parts of the world the game is most frequently played, whether there are any geographical or cultural aspects in terms of games frequency etc.

2 Data Overview and Hypotheses

Since the beginning of 2008 the results of all games among rated chess players have been published (with additional information) on the FIDE website http://www.fide.com/. While collecting the data, we ran into some problems. There are different players with the same name. Some players passed away during the analysis period and they are no longer in the FIDE database of players. For an unknown reason FIDE does not publish ratings for players from quite several countries. Harvesting from the Internet took time, because there were some interruptions in the connection and there were also some various minor typing errors. Last but not least, each game is listed twice. On the present FIDE’s players list
there are over a quarter million chess players from 167 countries and among them 92731
players were active during the analysis period. The word *active* in our context means
that they played at least one game from January 2008 to September 2010. They actually
played over 1.6 million games. Some properties of players and games are described in
this section.

Chess is one of the rare sporting events where men and women can compete against
each other. The vast majority of chess players in our database, over 92 % are male players.
Moreover, just 21 of the world’s 1170 (approximately 1.8 %) active chess grandmasters
are women. Chabris and Glickman (2006) noted that the lack of women who enter chess
competitions at the lowest levels is the main reason for the male domination of elite chess.
The proportion of men and women in their dataset is very similar to the one in our study
even though they studied the data from the US Chess Federation (USFC) exclusively.

Many years ago Elo realised that measurement of chess activity and interest is hardly
simple (Elo, 1978). Our advantage in comparison with his study is a large quantity of
(reliable) data. Analysing countries for which individual chess players play, we can con-
clude that the highest number of active competitors are from Germany, Spain, Russia and
France (left side of Figure 1). All four countries have a great chess tradition, especially
Germany has invested large amounts of funds in order to popularise the game of chess
in the last few years. In 2008 they hosted the Chess Olympiad (Dresden) and the World
Chess Championship match between the Russian Vladimir Kramnik and the Indian Vis-
nawathan Anand (Bonn). The strongest national chess league Bundesliga is also played
in Germany.

To obtain a more objective view, we divided the number of active players in each
country by the number of inhabitants in that country respectively. The situation changed
considerably, at the top we can find the Faroe Islands and Iceland (right side of Figure
1). ‘Iceland is an anomaly and a marvel’, a chess devotee visiting this country once
wrote. There is even one small island on Iceland, called Grimsey, where all men, women
and children are chess players. Climate, economic and socio-cultural circumstances on
Iceland do not favour many of the distracting western activities, but they do favour chess
(see Elo, 1978). The Faroe Islands are in the neighbourhood of Iceland, so they have a
similar sociodemographic situation with less inhabitants and this explains a lot.

The Elo Chess Rating System, named by Elo, is one of the most famous and used
rating systems. It was adopted by FIDE in 1970. The system is described in details by
the author in Elo (1978) and has many practical uses. For pairing purposes in tourna-
mements, it is in the best interests of organisers to avoid pairing the best players in early
rounds. Ratings can be used as a qualifying system for elite tournaments or events, to
complete national Olympiad teams or any team in general, for sectioning tournaments in
a specific range, where organisers allow only players of a specified rating range to com-
pete etc. Glickman and Jones (1999) considered it the most useful service of the rating
system because it allows competitors at all levels to monitor their progress. Youngsters
are interested in their rating progress as they are becoming better players, and the others
are also often interested in it due to comparison with their main rivals.

In Figure 2 the distribution of Elo rating points (or just rating points) among active
chess players is presented. Players with zero rating points are omitted because their first
rating points have not been calculated yet. We can observe that rating points are almost
normally distributed with a mean of around 2000 rating points. Players with 2700 rating points or more are usually called elite grandmasters. On the September 2010 rating list there were 39 players in this class. Norwegian Magnus Carlsen was on the top with 2826 rating points. To compare, the highest Elo rating ever achieved was 2851 rating points by Russian Garry Kasparow on the July 1999 and January 2000 rating lists.

In combinatorial game theory chess is considered as a two player perfect information partisan game where two players have different options for their moves. However, in reality chess is too complex to find an optimal strategy and it is not known how to play chess perfectly (Berlekamp, Conway, and Guy, 1982). A player with white pieces starts a chess game and for that reason has some advantage over a player with black pieces. Indeed, in our database approximately 54.4 % of all decided games were won by white. In Elo (1978) and Glickman (1995) slightly higher percentages were obtained, because not all games over an observed period were available. In the past (prior to January 2008) only the game results of better chess players and important games were published. It is known that the advantage of white pieces increases with the strength of both players (see Glickman, 1995) and this should be the main reason for differences.

Using the above described data and our knowledge of the current situation in the world of chess we developed three main hypotheses.

Figure 1: Presentation of a log number of all active chess players by country (left) and a normalised version (right).
Hypothesis 1. The best chess players in the world almost exclusively play against each other.

Firstly, the best players in the world mostly play against each other in order to maintain their high level of rating points. In some way, they form a very close society. They mostly play in the closed so-called Berger tournaments and very rarely attend open Swiss-type tournaments. To play in open tournaments it is very risky for world class chess players and it is a great potential threat for a loss of their rating points. In this computer era even a lower ranked chess player can be well prepared to play against the best players. In such a case everything is possible, not only chess knowledge and game understanding are decisive.

Hypothesis 2. The number of games played dependents on geographical position of countries. This dependence is inversely proportional to the strength of the players.

We assume that lower ranked players mostly play with opponents from the same country or some neighbouring countries due to financial and time constraints. But this is not the case with better players or even the world’s best players.

Hypothesis 3. An advantage of white pieces should be taken into account when the strength of a chess player is calculated.

The third hypothesis is about the Elo rating system. We have already seen that playing a game of chess with white pieces is a considerable advantage, but this has not been taken into account in the current method of calculating rating points so far and we think it should be.
3 Network(s) of Games Played

Basically, a network consists of set of actors (represented by vertices) and relations (represented by ties or edges) between them (see Wasserman and Faust, 1994). Edges may be directed or undirected and weighted or unweighted. A selection of our network is very natural, actors are chess players in our database and two players are related (connected with a tie) if they played at least one game between each other over the observed period. The obtained network is undirected and weighted by the number of games played between two opponents. We removed vertices with a zero degree (representing players not playing a single game) and the number of vertices decreases to 92731. The resulting network consists of 19 weak components, one huge weak component with 92677 vertices, one weak component with 12 vertices and other weak components with 5 vertices or less.

Figure 3 shows the degree distribution for the obtained network in a log-log scale. The degree starts very high and remains the same until it reaches 10, then the number of players decreases very rapidly. Five players played over 500 games (approximately one ranked game every two days): 625-Dimitar Marholev (BUL), 600-Pavel Sevostianov (RUS), 574-Milan Mrdja (CRO), 516-Viesturs Meijers (LAT) and 509-Mark Lyell (ENG). 500 ranked games are quite a lot if we consider that the average game lasts about four hours and all of those games should be official, played at tournaments or similar events. Therefore it is not surprising that only one of the listed players is a grandmaster (Meijers), but none of them is a world class chess player.

![Figure 3: Number of active chess players by degree in log-log scale.](image)

We are more interested in whether there is a correlation between the number of played games and the number of (different) opponents. In view of this, we define the same
opponents index (SOI) as:

\[ \text{SOI} = 1 - \left( \frac{a}{b} \right)^2, \]

where \( a \) is the number of different opponents and \( b \) is the number of games, both over the observed period. From (1) directly follows \( \text{SOI} \in [0, 1] \) and the higher the value of player’s SOI the more the player is playing against the same opponents. For example, the current number one, Magnus Carlsen, played 189 games against 84 different opponents. On the other hand Dimitar Marholev, who played the highest number of games in our database, played 625 games against 555 different opponents. Carlsen’s SOI is 0.80, Marholev’s 0.21 respectively.

Figure 4 represents the distribution of SOI by rating points and was obtained in the following way:
- we clustered all rated active chess players with non-zero rating points into 16 groups, with the width of 100 points each,
- we calculated the mean rating of each group,
- we calculated the mean SOI of each group,
- points are plotted as ordered pairs of the last two calculations.

![Figure 4: Distribution of the same opponents index by rating points.](image)

The SOI is very low in the group of players with rating points between 1200 and 1300 following an almost constant low shape among the players with rating points below 2100. The players with rating points below 2100 rating points mostly play against different opponents - on average their SOI is very low. From here on it starts to grow exponentially and from this point it can be interpreted as better players play repeatedly against almost the same rivals. Therefore, the best players mostly play against each other thus confirming our first hypothesis.
### 3.1 Denser Parts of a Network

With cores (in our case valued cores or p-cores) we can identify clusters of vertices in a network that are tightly connected (see de Nooy, Mrvar, and Batagelj, 2005). The 57 p-core is the highest valued core in our undirected network of played games. Following our close review of all high valued p-cores we decided to analyse 52 p-core in more details. First, we deleted all valued cores lower than the 52 p-core to obtain Figure 5. Each player in this subnetwork played a minimum of 52 games with other players in this 52 p-core. Vertices are coloured by country partitions. A different colour of two vertices consequently means a different country of origin. The Kamada-Kawai algorithm (see de Nooy et al., 2005) was used to determine the locations of the vertices.

![Figure 5: 52 valued core.](image)

The induced subnetwork consists of one component, but it is obviously broken into three parts. The largest part is composed of Indian chess players (yellow coloured vertices) and the top Bangladeshi player Rahman Ziaur. India plays an important role in the history and evolution of the chess game and even the recent world champion Viswanathan Anand comes from India (we can find him in the upper part of Figure 5). So, it is not surprising to see Indian chess players playing so many games against each other.

The view on the right side of Figure 5 is more surprising, almost all of the players are Ecuadorians (the only exception is the Cuban player Medina Miguel). In September 2010 Ecuador was placed only in 63rd rank in terms of the average rating points of top 10 players on the FIDE country rankings and it had only one grandmaster (India on the other hand was 4th on the same country ranking list and it had 22 grandmasters, Russia was convincingly leading with a total of 201 grandmasters). We can notice the special position of Ecuadorian player Macias Murillo Bryan, who joined the 'Indian' and 'Ecuadorian' parts together.
Arguably, the most interesting part is the top part of the analysed valued core. It consists of 20 elite chess players. On the September 2010 rating list all of them had over 2700 rating points. The number of countries where players originate from is much more varied than in the other two parts. There are six Russian players, two Azerbaijani and Ukrainians and one Chinese, Indian, Bulgarian, Hungarian, Armenian, Norwegian, Spanish, French, American and Israeli player. It is obvious for the world’s best chess players mutual distance does not represent a major barrier and they tend to play a lot against each other. The strongest link between the best players at the top of Figure 5 and the Indians is made by the established Indian grandmaster Surya Shekhar Ganguly. This valued core presentation confirms our first two hypotheses about the best players and geographical dependence.

3.2 Subnetwork of the World’s Best Chess Players

FIDE issues a ratings list once every two months. However, before July 2009 the ratings list was published once every three months. There were only 17 chess players with their rating points of 2700 or more on all published rating lists throughout the analysed period and this really makes them the best of the best. Their names can be seen in Figure 6 where all active players actually take part. Figure 6 was obtained in the following way. First, all other players apart from 17 best players were shrunk into five clusters (red coloured vertices), according to their rating points, named as: 0+, 2000+, 2200+, 2400+, 2600+. Players with rating points between 0 and 1999 are in the 0+ cluster, players with rating points between 2000 and 2199 are in the 2000+ cluster and so on. We deleted all the loops and edges between shrunken vertices and also between the 17 best players. The only remaining edges were between the 17 best players and the shrunken vertices. At the end we normalized the remaining values on edges by the number of all games played by each of the 17 best players. A darker colour of the edges means that a larger portion of games was played against players in a shrunken vertex.

![Figure 6: Network of games between 17 best chess players and others.](image)

The best players in Figure 6 are mainly divided into four groups. Kramnik is left
alone, playing only with the 2600+ group. In almost three years he never played a single rated game against a chess player with rating points lower than 2600! The right group, consisting of five players, played only against the 2600+ and 2400+ clusters and the central group played against 2600+, 2400+ and 2200+. Only two players played also against the 2000+ cluster, but actually nobody played against the players rated under 2000 rating points, which is the mean rating of all active players. We can also conclude that players of each group have the strongest links to the 2600+ cluster. This presentation is a strong approval of our first hypothesis.

3.3 Games Played Between Countries

We want to examine how often a game of chess is played among players from different countries to get some form of a global view. We expect that chess tradition and recent political changes in some Eastern European countries could play an important role.

First, we shrunk all of the players from each country into one vertex and then we deleted all of the loops - games among players from the same country. In this way we obtained 167 vertices and we kept the number of edges between any pair of vertices (countries) in a shrunken network if a minimum of 100 games were played between those two countries. Next, we performed a kind of normalisation; we divided the number of games between two countries by the product of the square root of the number of active players in those two countries. At the end we classified all of the remaining countries into six clusters using the Ward method with a Corrected Euclidean distance (see Doreian, Ferligoj, and Batagelj, 2005) to obtain a clearer picture. In such a way, the matrix in Figure 7 was obtained. Squares representing the normalised number of games played between two countries are darkened from white to black in four levels. There are some grey and even black spots in this matrix and we will look into those patterns in more detail.

The largest dark spot on the main diagonal illustrates a huge portion of games played among the countries in the third cluster (from the bottom up). In this cluster we can find seven Asian countries: Malaysia, Myanmar, China, Vietnam, the Philippines, Singapore and Indonesia. Obviously, players from these countries play against each other a lot and also against other Asian country — India, placed in the second cluster (from the bottom up).

The bottom right-hand corner forms a cluster of central European countries, which is well known in the chess world. It consists of the following countries: Hungary, the Netherlands, the Czech Republic, Poland, France, Italy, Spain, England, Russia, Germany and Ukraine. All these countries have a rich chess tradition. They have raised many world famous grandmasters and the most world chess champions have come from this part of the world. Players in that cluster mainly played each other (they form a complete block) and against players from the second cluster which consists of some European countries (ex-Yugoslavian countries, some ex-Soviet Union countries, etc.), Israel, India, the USA and others.

The fourth cluster is also interesting. There we can find countries from Central America and South America: Venezuela, Mexico, Cuba, Colombia, Ecuador, Bolivia, Peru, Brazil, Argentina, Paraguay and Uruguay. In addition to playing against each other, they
also played against the players from the USA and Spain. Against USA due to its geographical proximity, against the latter probably due to a common culture and language (Latino countries).

Figure 7 also exposes pairs of neighbouring countries whose players often played against each other: Azerbaijan and Georgia, the Former Yugoslav Republic of Macedonia and Bulgaria, Lithuania and Latvia, Switzerland and Liechtenstein, Uruguay and Paraguay, Costa Rica and Nicaragua, the United Arab Emirates and Syria, Uzbekistan and Kazakhstan, Australia and New Zealand, Egypt and Jordan, Bangladesh and Nepal. At the top of Figure 7 there is a cluster of over 60 countries, mostly small ones and in terms of chess underdeveloped countries. For those countries we can say that their players mainly played against their countrymen. But remember, the main diagonal is empty due to deleted loops which represent the games played among players from the same country. We can conclude that the geographical position of a country has a major impact on
the frequency of chess games played among countries. Chess players from neighbouring countries played each other more often.

### 3.4 Elo Rating System

Elo based his theory on two key assumptions (Ross, 2007). Firstly, the strength of each chess player is a normally distributed random variable with some unknown mean (players’ strength) and known variance. Moreover, he assumed that the variances of strength of all players are the same. Under this two assumptions every player’s distribution of strength follows a normal distribution and therefore having the same shape centered at a different value depending on player’s overall ability.

The difference of two normally distributed random variables is again a normally distributed random variable, where the mean of the latter is the difference of means of initial distributions and the variance is the sum of both variances. For ease of computation, the cumulative distribution function of a normal distribution was replaced by its approximation — a logistic function. Consequently, in a game between players with a strength $R_W$ (strength of a player with white pieces) and $R_B$ (strength of a player with black pieces) the expected score of the game for a player with white pieces is assumed to be:

$$E = \frac{1}{1 + 10^{-(R_W - R_B)/400}}.$$  \hfill (2)

Actual score of the game is 1 if player $W$ wins, $1/2$ if the game ended in a draw and 0 if player $B$ wins. As a statistical model (2) is often termed as the winning expectancy formula (see Glickman and Jones, 1999). Formula 2 can also be interpreted as an extended type of Bradley-Terry model (Bradley and Terry, 1952) where an additional result (a draw) is possible. In Figure 8 a logistic curve representing equation (2) is presented in black colour. Obviously, this curve is considered point symmetrical (over a point $(0, 1/2)$), which means that it does not take into account differences between white and black pieces in the expected outcome of the game.

In the past some authors already tried to improve the winning expectancy formula. The most prominent were Glickman and Jones (1999) and Sonas (2002), the latter even proposed a linear formula. Due to lack of information Glickman and Jones (1999) did not incorporate colour advantage although they were aware of it. Their formula $1/(1 + 10^{-(R_W - R_B)/561})$ is presented with green colour in Figure 8 and we can easily notice that model is point symmetrical over a point $(0, 1/2)$. On the other hand Sonas (2002) incorporated colour advantage, but the linear model $0.541767 + 0.001164 \cdot (R_W - R_B)$ (in Figure 8 presented with blue colour) is valid only on minor interval because the expected result of a chess game should always be in closed interval between 0 and 1.

In accordance with (2) we postulated a model for the average observed score as $1/(1 + 10^{-(\Delta r + \alpha)/400})$, where $\Delta r$ is the difference in ratings of both players ($\Delta r = R_W - R_B$) and $\alpha$ is an unknown scaling parameter to be inferred from our data. With the least square method our prediction of a game’s outcome based on published games is $1/(1 + 10^{-(\Delta r + 28.44)/400})$. The fitted model is drawn in Figure 8 as a red dotted line. The expected result of a player with white pieces in (2) is obviously underestimated, confirming our last hypothesis.
To clarify the use of the expected result we should add that the expected result in a game of chess is usually denoted by $W_e$ and it is one of the key parameters for the calculation of new rating points of an individual chess player. The latter is done by using a formula:

$$R_{post} = R_{pre} + K \cdot (W_a - W_e),$$

where $R_{pre}$ and $R_{post}$ are player’s rating points before and after the game. $K$ is a coefficient based on player’s current rating points (it is known prior to the game) and $W_a$ is an achieved result of the game. All parameters on the right side of equation (3) are known prior to the game or when the game ends, the key question is how to determine the expected result $W_e$. Some propositions were introduced in this section and are plotted in Figure 8.

4 Conclusion

This study analysed the rated chess games played from January 2008 to September 2010 and published on the FIDE website. Our first assumption was that the world’s best chess players formed a very closed cohort. Network analysis approach findings, valued core presentation and a shrunk network of best players, led us to believe that best players do indeed very carefully select the tournaments and events in which they participate and therefore they indirectly even choose their opponents. The consequences are a slower and aggravating progression of young and ambitious chess players who can hardly get a chance to play against the best players. In this context, the privilege is held by countries.
with a long and great chess tradition, where the vast majority of the strongest tournaments are organized. Even some of the best players are dissatisfied with the current situation as illustrated by the latest example: world number one Magnus Carlsen withdrew from the upcoming World Chess Championship cycle at the beginning of November 2010 (see Chess.com, 2010).

Globally, countries were divided into six clusters in view of the normalized number of games played between them. A common characteristic of the individual clusters was primarily geographic proximity as well as in some cases an associated chess tradition and culture. Asian countries stand out significantly because in this part of the world chess is developing very quickly and continues to make considerable progress. After all, both world chess champions in female and absolute categories are from Asia.

Playing a game of chess with white pieces is an advantage which was verified also in our database. In the current rating calculations this is not considered yet. Consistent with the results of Glickman and Jones (1999) and Sonas (2002) we showed that this advantage should be taken into account in the future rating calculations.

**Computational and visualization details**

The programs for collecting the data from the FIDE website and producing networks were written in R (see R Development Core Team, 2010). For analysis of networks we used the program Pajek (see Batagelj and Mrvar, 1996-2010).

**References**


Authors’ addresses:

Kristijan Breznik
International School for Social and Business Studies
Mariborska cesta 7
3000 Celje
Slovenia

Vladimir Batagelj
University of Ljubljana
Faculty of Mathematics and Physics
Jadranska 19
1000 Ljubljana
Slovenia