CS552
Chess Network Analysis

Abinash Acharya (11840050)

Indian Institute of Technology Bhilai , abinasha@iitbhilai.ac.in

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 . This paper analyses all of the rated chess games of the top 1000 rated players published on the website https://365chess.com from January 2010 to March 2021 with network analytic methods. We discuss relations between player-player interactions and its correlation with their federations and ratings

Keywords: Chess · FIDE · Rating · Centrality · Community · Degree · Distribution

1 Introduction

Chess is an ancient game. In the last hundred years, chess has been closely connected to the world chess federation FIDE. It was founded in 1924 in Paris and recognised by the International Olympic Committee as an International Sports Federation in 1999.
FIDE is an international organization based in Switzerland that connects the various national chess federations and acts as the governing body of international chess competition. It is also its responsibility to calculate the rating points — a measure of the strength of each chess player.
FIDE awards titles for achievement in competitive play, such as the Grandmaster title. It also awards titles to composers and solvers of chess problems and studies.
Chess has made a very important contribution to strength measuring of players in various sports.
In this paper we will analyse all the rated games of the top 1000 currently active chess players featured in https://365chess.com played from January 2010 to March 2021 with network analytic methods. A more detailed description of players and game properties will follow in the next section.
I believe that 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 the correlation between the players and other various player attributes.

2 Data Collection

I picked the top 1000 active players by their current rating from https://365chess.com. Some important things to note are :

- Active Players :- These are those players who have at least played a single game on or after 2019 . This was done to filter out all the older players who had retired long ago

- All games played by the players from January 2010 to March 2021 were extracted from https://365chess.com and games consisting opponents from the 1000 nodes were kept and others were discarded
- Following are some of the statistics about the data extracted:
  No. of Players - 1000
  No. of matches - 1,24,665
- The fields extracted (for players) are :
  Rank :- Rank within the 1000 players in terms of higher rating
  Gender :- Male/Female
  Name
  Rating :- Current FIDE rating
  Title :- Any title received
  FIDE ID :- ID assigned to the player by FIDE
  Nationality
  Number of games :- Total Classic rated games played by the individual
  Year of Birth

Players

	Id	Gender	Rating	Title	Nationality	Games	Rank	Fide Id	Year of Birth
0	Magnus Carlsen	Male	2847	Grand Master	Norway	2835	1	1503014	1990
1	Fabiano Caruana	Male	2820	Grand Master	United States	2326	2	2020009	1992
2	Liren Ding	Male	2791	Grand Master	China	1507	3	8603677	1992
3	Ian Nepomniachtchi	Male	2789	Grand Master	Russian Federation	2081	4	4168119	1990
4	Levon Aronian	Male	2781	Grand Master	Armenia	3024	5	13300474	1982
...	...	...	...	...	...	...	...	...	...
995	Vadim Malakhatko	Male	2468	Grand Master	Belgium	1698	996	14104202	1977
996	Alexander Bagrationi	Male	2468	Grand Master	Israel	72	997	14104954	1990
997	Tim Janzelj	Male	2468	International Master	Slovenia	157	998	14607913	1994
998	Ogulcan Kanmazalp	Male	2468	International Master	Turkey	411	999	6304990	1992
999	Evgeny Gleizerov	Male	2468	Grand Master	Russian Federation	1443	1000	4101332	1963

1000 rows × 9 columns

Figure 1: A Snippet of the Extracted players' data

Matches

	White	ELO_White	Black	ELO_Black	Outcome	Moves	Opening	Year	Tournament
0	Magnus Carlsen	2847.0	Wesley So	2770.0	Draw	42	B52	2021	Magnus Carlsen Inv KO
1	Wesley So	2770.0	Magnus Carlsen	2847.0	Loss	38	E20	2021	Magnus Carlsen Inv KO
2	Wesley So	2770.0	Magnus Carlsen	2847.0	Draw	44	A20	2021	Magnus Carlsen Inv KO
3	Magnus Carlsen	2847.0	Wesley So	2770.0	Draw	38	C46	2021	Magnus Carlsen Inv KO
4	Magnus Carlsen	2847.0	Wesley So	2770.0	Win	23	C46	2021	Magnus Carlsen Inv KO
...	...	...	...	...	...	...	...	...	...
124660	Renato R. Quintiliano Pinto	2430.0	Jose Fernando Cubas	2425.0	Draw	35	D10	2016	3rd Floripa Masters 2016
124661	Renato R. Quintiliano Pinto	2422.0	Jose Fernando Cubas	2455.0	Win	6	A56	2015	VIII Hebraica GM 2015
124662	Renato R. Quintiliano Pinto	2364.0	Jose Fernando Cubas	2459.0	Draw	12	D38	2015	Floripa Masters 2015
124663	Ogulcan Kanmazalp	2439.0	Vadim Malakhatko	2488.0	Draw	10	C07	2017	ISEM URLA GM 2017
124664	Vadim Malakhatko	2488.0	Ogulcan Kanmazalp	2439.0	Win	36	A40	2017	ISEM URLA GM 2017

124665 rows × 9 columns

Figure 2: A Snippet of the Extracted Matches data

3 Network Creation

A temporal directed network with the following configuration was created

- Nodes - Players
- Edges - An edge from player A to player B represents player A has won more matches against player B (than B against A) out of all extracted matches
- Node Attributes - Name (Id) ,Rank,Gender,Rating,Title,FIDE Id,Nationality, Number of games , Year of Birth
- Edge Attributes -
  1. Number of matches :- This represents matches played between the 2 players within the time period
  2. Win Difference :- No. of matches won by source Node against the target - No. of matches won by target Node against the source
  3. Adjusted Win Difference :- (Win Difference)/No. Of Matches
- Each Edge has a timestamp associated with it. The timestamps range from 2010 to 2021
- Each Edge corresponds to games played cumulatively from 2010 to the timestamp associated.
- Alternatively, the timestamp can be represented as a one-year interval as well. For example, 2015 can be represented as 2015-2016 as a 1-year interval

EdgeTable

Source	Target	Weight	Matches	time_start	time_end	Interval
Magnus Carlsen	Nigel D Short	1	2	2010	2011	<[2010,2011]>
Vladimir Kramnik	Magnus Carlsen	2	8	2010	2011	<[2010,2011]>
Magnus Carlsen	David W L Howell	1	1	2010	2011	<[2010,2011]>
Magnus Carlsen	Hikaru Nakamura	1	4	2010	2011	<[2010,2011]>
Viswanathan Anand	Magnus Carlsen	1	10	2010	2011	<[2010,2011]>
Michael Adams	Magnus Carlsen	0	2	2010	2011	<[2010,2011]>
Magnus Carlsen	Michael Adams	0	2	2010	2011	<[2010,2011]>
Luke J McShane	Magnus Carlsen	1	1	2010	2011	<[2010,2011]>
Magnus Carlsen	Dmitry Andreikin	2	2	2010	2011	<[2010,2011]>
Rauf Mamedov	Magnus Carlsen	0	2	2010	2011	<[2010,2011]>
Magnus Carlsen	Rauf Mamedov	0	2	2010	2011	<[2010,2011]>
Teimour Radjabov	Magnus Carlsen	0	4	2010	2011	<[2010,2011]>
Magnus Carlsen	Teimour Radjabov	0	4	2010	2011	<[2010,2011]>
Magnus Carlsen	Boris Grachev	0	2	2010	2011	<[2010,2011]>
Boris Grachev	Magnus Carlsen	0	2	2010	2011	<[2010,2011]>
Magnus Carlsen	Boris Savchenko	0	2	2010	2011	<[2010,2011]>
Boris Savchenko	Magnus Carlsen	0	2	2010	2011	<[2010,2011]>
Magnus Carlsen	Ian Nepomniachtchi	0	2	2010	2011	<[2010,2011]>
Ian Nepomniachtchi	Magnus Carlsen	0	2	2010	2011	<[2010,2011]>
Magnus Carlsen	Shakhriyar Mamedyarov	1	2	2010	2011	<[2010,2011]>

Figure 3: Snippet of the Edge Attributes

4 Analysis

4.1 Statistics

- No of nodes – 1000

- No of Edges – 39969 (Throughout the timeline)
- Average Degree – 39.969
- Average Weighted Degree – 50.224
- Network Diameter – 6
- Average Path Length – 2.41
- Average Clustering Coefficient - 0.137
- No. of Strongly Connected Components - 14
- No. of Weakly Connected Components - 4

4.2 Node Attributes

Nationality

The distribution of players in terms of nationality is as follows :

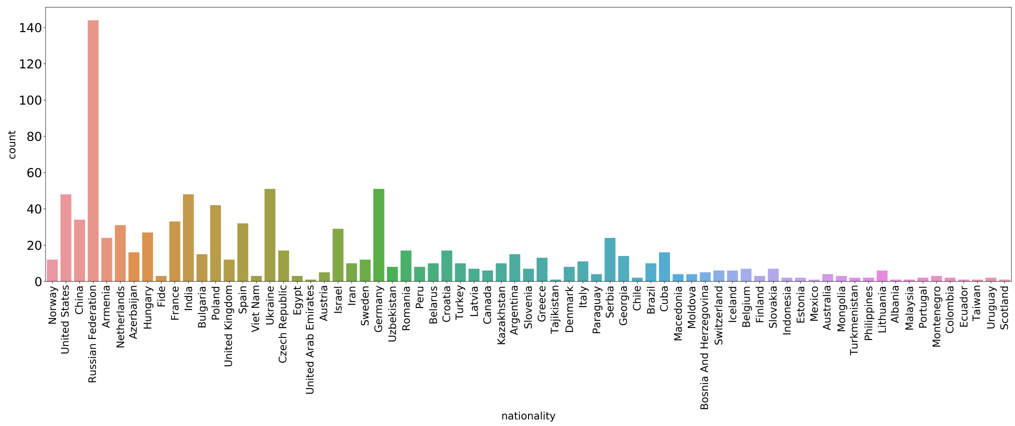

Figure 4: Nation/Federation Wise Player Distribution

As we can see most players are from the Russian Federation, followed by European countries and India.

Title

The Title wise distribution of players is as follows :

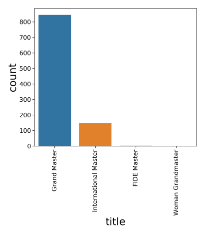

Figure 5: Title Wise Player Distribution

Most players are Grand masters. Several players are International masters. There are also 2 FIDE masters and 1 Woman grandmaster.

Gender

The gender wise distribution of players is as follows :

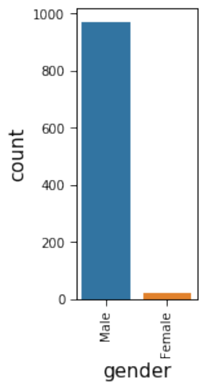

Figure 6: Gender Wise Player Distribution

It is evident from the plot that most players are males.

Ratings

The rating wise distribution of players is as follows :

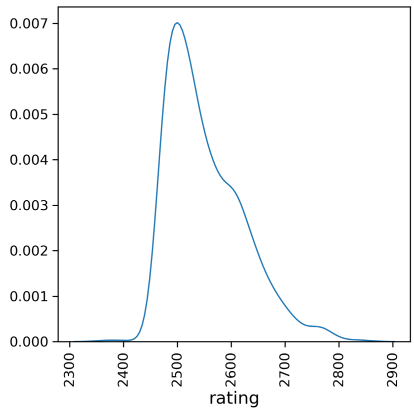

Figure 7: Rating Distribution of players

Several ranges for ratings were defined and the distribution is as follows :

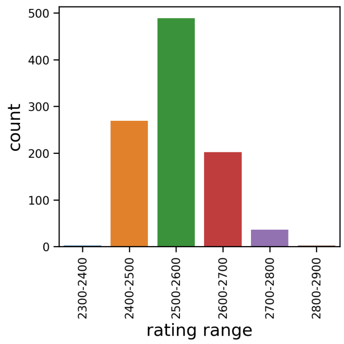

Figure 8: Category Wise Distribution of players according to defined range of ratings

From Figure 8, it is evident that most players belong to the range of 2500-2600 in terms of rating and there are very few players who are rated 2700 or higher.

4.3 Degree Distribution

The log-log plot (Green - In-degree , Blue - Out-degree) for Degree Distribution (Degree vs Frequency) is given as follows for 3 different timestamps - 2010 (Initial), 2015 (Mid) , 2021 (End).

- 2010

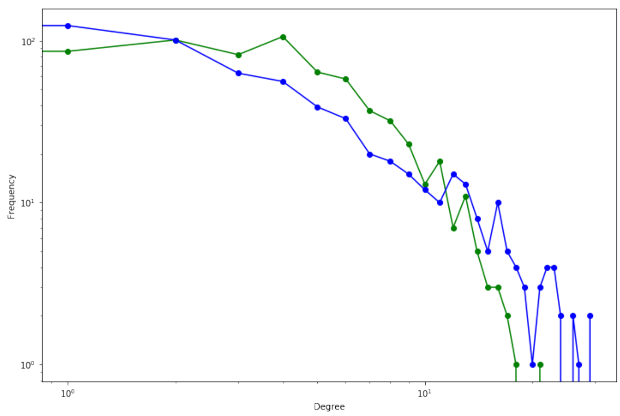

Figure 9: Out-Degree vs Frequency(Blue Line) and In-degree vs Frequency(Green Line) Plot for nodes in 2010

- 2015

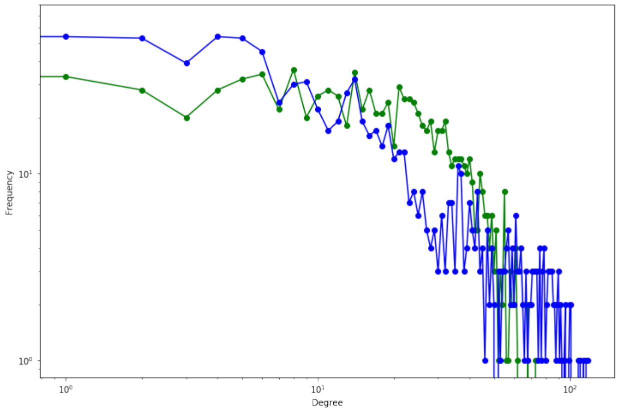

Figure 10: Out-Degree vs Frequency(Blue Line) and In-degree vs Frequency(Green Line) Plot for nodes from 2010-2015

- 2021

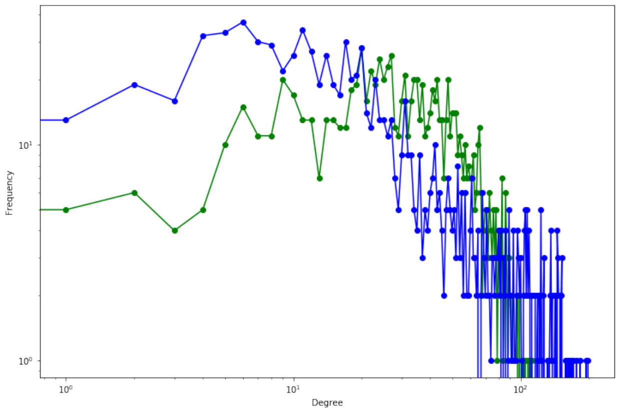

Figure 11: Out-Degree vs Frequency(Blue Line) and In-degree vs Frequency(Green Line) Plot for nodes from 2010-2021

Observations and Inferences

1. Approximated Power Exponent, $\gamma = 1.25$ (For 2021)
2. Approximated Power Exponent for out-degree, $\gamma_{out} = 1.323$ (for 2021)
3. Approximated Power Exponent for in-degree, $\gamma_{in} = 1.29$ (for 2021)
4. In 2010, however,$\gamma = 1.75$
5. Scale-free nature can be seen initially, but with more games being played each year, the network becomes densely connected losing its scale-free nature
6. In 2021, The graph is not shaped like a power-law plot. Rather it has a flat maxima in the middle (after 10 degree mark). This suggests most nodes are moderately connected. There are very few nodes with extreme degrees.

4.4 Centrality Measures

4.4.1 Degree Centrality

Degree Centrality measures (for Out-Degree) were done on the following circumstances:-

- Without Weights
- Using Win difference as weight
- Using Number of Matches played as weight
- Using Win Difference / No. of Matches as weight

Top 5 nodes in terms of out-degree centrality for each of the above edge weights is as follows (for timestamp 2021) :
Without Weights - Rating

1. David Navara - 2697
2. Anton Korobov - 2683
3. Maxim Matlakov - 2688
4. Maxime Vachier Lagrave - 2758
5. Ian Nepomniachtchi – 2789

Using Win difference as weight - Rating

1. Magnus Carlsen - 2847
2. Hikaru Nakamura - 2736
3. Maxime Vachier Lagrave - 2758
4. Ian Nepomniachtchi - 2789
5. David Navara - 2697

Using Number of Matches played as weight - Rating

1. Magnus Carlsen - 2847
2. Hikaru Nakamura - 2736
3. Maxime Vachier Lagrave – 2758
4. Levon Aronian - 2781
5. Ian Nepomniachtchi - 2789

Using (WIn Difference / No.of Matches) as weight - Rating

1. David Navara - 2697
2. Igor Kovalenko - 2643
3. Anton Korobov - 2683
4. Arkadij Naiditsch - 2649
5. Rauf Mamedov - 2654

4.4.2 Betweenness Centrality

Although this is not a flow network, we tried to find if the nodes having high betweenness centrality bear any significance.
Top 5 nodes in terms of betweenness centrality (without considering edge weights) is as follows (for timestamp 2021) :

1. Arkadij Naiditsch - 2649
2. Igor Kovalenko - 2643
3. Romain Edouard - 2611
4. Alexandr Fier - 2565
5. Marin Bosiocic - 2601

4.4.3 Eigenvector Centrality

Top 5 nodes in terms of eigenvector centrality for each of the below edge weights is as follows (for timestamp 2021) :

<u>Using Win difference as weight - Rating</u>

1. Boris Savchenko - 2552
2. Pavel Potapov - 2481
3. Mikhail Al. Antipov - 2609
4. Vahap Sanal - 2571
5. Dmitry Bocharov - 2533

<u>Using (WIn Difference / No.of Matches) as weight - Rating</u>

1. Vahap Sanal - 2571
2. Valentin Dragnev - 2564
3. Mikhail Al. Antipov - 2609
4. Pavel V Tregubov - 2560
5. Tal Baron - 2522

<u>**Observations and Inferences**</u>

- **Unexpectred results for Eigenvector Centrality** - Some moderately rated players have higher eigenvector centrality than some of the highest rated players.
- **Possible Explanation** – These players are not expected to beat high rated players but since they have done so, they have higher eigenvector centrality.
- **Inference** - These players have performed better than expectations.
- Some players like Arkadij Naiditsch and Igor Kovalenko have a high betweenness centrality as well as a high degree centrality with (Win difference/Matches) as weight.This might be because these players are moderately rated but have defeated some high rated players. Although they have played few matches with high rated players but have an upper hand over them in that small sample size.

4.5 Community Detection

- 9 communities were detected after applying Louvain method in Gephi
- The modularity for the partition obtained was 0.238
- These communities have been visualized in Gephi and the nodes have been colored according to the community they belong to
- Edge weight corresponding to win difference has been taken into account while forming communities

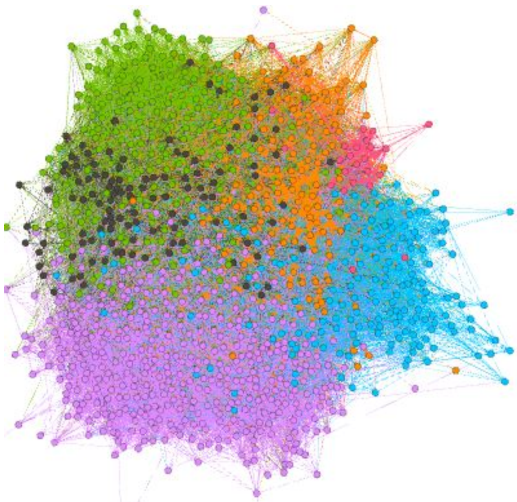

Figure 12: Communities Obtained by Louvain Algorithm

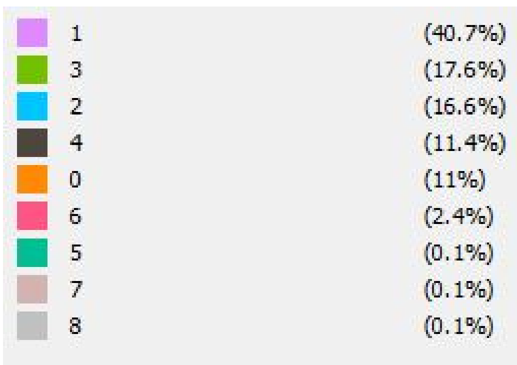

Figure 13: Color Labels for Figure 12

The distribution of communities is as follows :

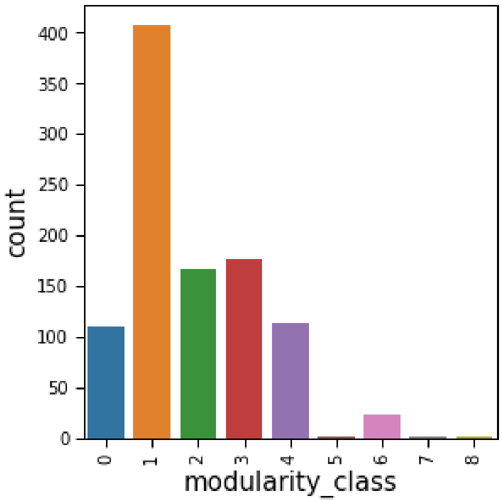

Figure 14: Community Size Distribution

Even though there are 9 communities, 3 of them are trivial consisting of a single player each. These are communities 5,7 and 8.
Community 1 is abundant and consists of the highest number of players.
I tried to find correlation between the communities and various node attributes.The results are given below

Correlation of Communities with Rating

I plotted the distribution of rating groups inside all the communities obtained. Some communities consist of a high majority of people from a certain rating range group as we can see in the plot below.

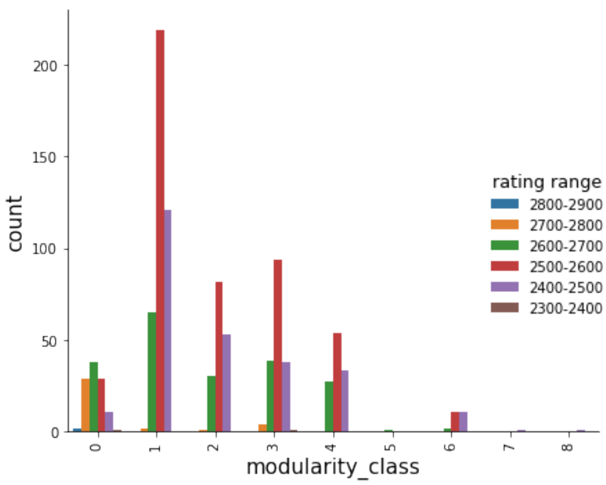

Figure 15: Rating composition inside Communities

We can see that Community 0 consists of a large fraction of individuals from the rating ranges of 2600-2700,2700-2800 and 2800-2900.
Out of these, players belonging to 2700-2800 and 2800-2900 are almost exclusive to Community 0. Therefore, this can be considered as the community of highest-rated players.

Community 1 ,on the other hand, consists of players mostly rated between 2400-2600 and a few between 2600-2700. So, this community is the group of moderately rated individuals. Other communities have a composition similar to that of Community 1.
The distribution of communities inside the rating groups has been shown below :

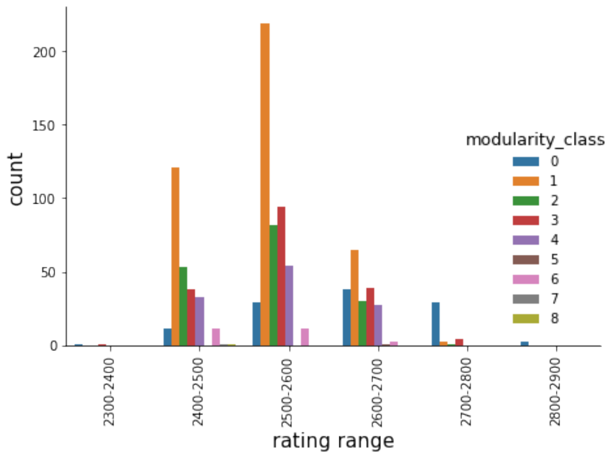

Figure 16: Community Wise Composition of Rating Groups

This plot again re-iterates that 2700-2900 players mostly belong to community 0 and players belonging to the 2400-2600 range form the community 1

Observations and Inferences

Highest rated players, especially those from 2700-2900 , play frequently against each other and moderately rated players between 2400-2600 play against each other and have rivalries.

Correlation of Communities with Nationality

Since there are 71 federations/nations which the players belong to, we will use only a subset of all the nations to study their correlation with obtained communities. Here, we have chosen the top 8 nations (sorting by number of players belonging to nations) namely, Russian Federation,Germany,Ukraine,United States,India,Poland,China,France.
About 50% of all the players belong to these 8 nations.

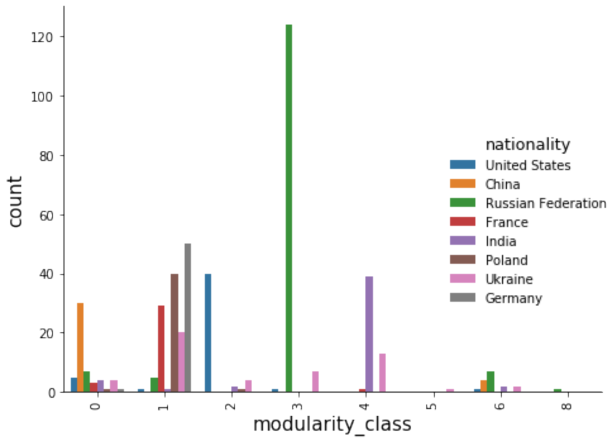

Figure 17: Nation Wise Distribution in communities

From figure 17, we can see that community 0 primarily consists of players from China,community 2 consists of players from the US, community 3 consists of Russian players and community 4 mostly consists of Indian players. Other communities have mixed nation-wise composition. Now, we will observe the community composition in various national groups/federations.

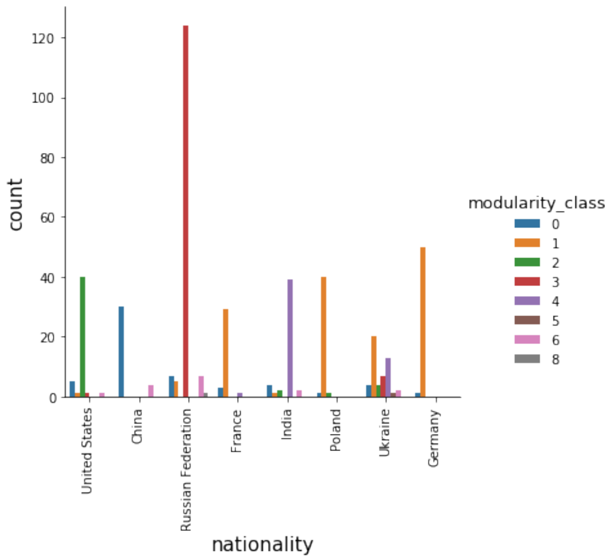

Figure 18: Community Wise Composition in players of various nationalities

From figure 18, it is clear that players of China are concentrated in community 0, players of the United States are concentrated in community 2, Russian players in community 3, Indian players in community 4 and other nations have players from various communities with community 1 being in majority since it is the largest community by size.

Observations and Inferences

There is a certain degree of correlation between the communities obtained and the nationalities of players. Players from India, US, Russia and China belong to distinct communities implying that these players face players within the nation frequently.

4.6 Assortativity Analysis

4.6.1 Degree Assortativity

The out-degree assortativity at each timestamp (from 2010 to 2021) was calculated. Here are the results :

Table 1: Degree Assortativity for timestamps between 2010-2021

Year	Degree Assortativity
2010	0.372
2011	0.325
2012	0.315
2013	0.317
2014	0.3
2015	0.278
2016	0.259
2017	0.26
2018	0.237
2019	0.226
2020	0.227
2021	0.226

As we can see degree assortativity is substantial in the beginning and starts to dwindle with time. This might be due to the sample size of matches increasing with each year and hence, higher rated players getting to play with lower rated players more.
In spite of that, degree assortivity is still reasonably high by 2021 implying that players tend to play matches with players having similar out-degree i.e. similar experience or performance (not necessarily similar rating).

4.6.2 Assortativity based on Nationality

The assortativity in the network with respect to nationality of players at each timestamp (from 2010 to 2021) was calculated. Here are the results :

Table 2: Assortativity with respect to nationality for timestamps between 2010-2021

Year	Assortativity w.r.t nationality
2010	0.169
2011	0.177
2012	0.167
2013	0.167
2014	0.162
2015	0.152
2016	0.147
2017	0.144
2018	0.139
2019	0.131
2020	0.131
2021	0.131

From Table 2, it can be seen that assortativity in terms of nationality also declines throughout the years, following a similar trend like degree assortativity.

Although assortativity w.r.t nationality is a bit lower than corrsponding degree assortativity, it is still fairly significant in magnitude. Hence, it can be concluded that players belonging to the same nationality play each other more often.

4.6.3 Assortativity based on Rating

The assortativity in the network with respect to rating of players at each timestamp (from 2010 to 2021) was calculated. Here are the results :

Table 3: Assortativity with respect to rating for timestamps between 2010-2021

Year	Assortativity w.r.t rating
2010	-0.001
2011	0.0
2012	-0.001
2013	-0.001
2014	-0.002
2015	-0.001
2016	-0.001
2017	-0.001
2018	-0.001
2019	-0.001
2020	-0.001
2021	-0.001

The assortativity in terms of rating is zero or negative throughout all the years suggesting that players don't only play with players of similar rating. Players play matches and beat opponents of different ratings.
I attempted to treat rating as a categorical variable rather than a continuous one to see if assortativity changes.
The assortativity in the network with respect to rating ranges of players, as shown in figure 8 at each timestamp (from 2010 to 2021) was calculated. Here are the results :

Table 4: Assortativity with respect to rating-ranges for timestamps between 2010-2021

Year	Assortativity w.r.t range of ratings
2010	0.068
2011	0.077
2012	0.067
2013	0.064
2014	0.063
2015	0.059
2016	0.055
2017	0.052
2018	0.046
2019	0.043
2020	0.042
2021	0.042

From Table 4, it is evident there there is some improvement in assortativity when ratings are replaced with rating ranges.

Although still the values are quite small in magnitude suggesting that there is very little correlation between the ratings of neighboring nodes.

5 Conclusion

We studied the network and analysed it using various methods such as the composition of node attributes, the degree distribution, the importance of nodes in terms of various centrality measures. We used community detection methods to find similarities of these communities with various other node attributes. We also analysed the assortativity of the network for degree and other attributes.

6 Future Work

We can use this study to make predictions on future matches by using the analysis.
Some other methods like machine learning algorithms can be used to use the network embedding to predict matches.
Additionally, further work can be done to study the network as time-series data (since it is a temporal network) and analyse it further.

References

[1] Kristijan Breznik, Vladimir Batagelj, *FIDE Chess Network*, Slovenia,2011,URL:https://pdfs.semanticscholar.org/d687/a3021357017d3f4b17de08efb8b73184cb4a.pdf.

[2] Mathieu Bastian , Sebastien Heymann,*Gephi : An Open Source Software for Exploring and Manipulating Networks* , URL : https://gephi.org/publications/gephi-bastian-feb09.pdf

[3] Haoran Wang, *Louvain Community Detection Algorithm*, 2020 , URL : Implementation of Louvain Community Detection Algorithm in Gephi

[4] *Chess Database*, 2021 , URL : https://365chess.com