

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

ANALÝZA ŠACHU AKO HERNEJ KOMPLEXNEJ SIETE

Bakalárska práca

2019

Pavol Freivolt

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

ANALÝZA ŠACHU AKO HERNEJ KOMPLEXNEJ SIETE

Bakalárska práca

Študijný program: Aplikovaná informatika

Študijný odbor: 2511 Aplikovaná informatika

Školiace pracovisko: Katedra aplikovanej informatiky

Školiteľ: Mgr. Peter Náther, PhD.

Bratislava, 2019

Pavol Freivolt



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Pavol Freivolt
Študijný program: aplikovaná informatika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor: aplikovaná informatika
Typ záverečnej práce: bakalárska
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Analýza šachu ako hernej komplexnej siete
Analysis of chess as a complex game network

Anotácia: .

Cieľ: Analýza a kategorizácia šachových pozícií, vybudovanie siete vytvorenej ťahmi a analýza tejto siete.

Literatúra: The game of go as a complex network, B. Georgeot , O. Giraud, CoRR vol. abs/1105.2470 2011
<http://arxiv.org/abs/1105.2470>

Vedúci: Mgr. Peter Náther, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: prof. Ing. Igor Farkaš, Dr.
Dátum zadania: 05.10.2017

Dátum schválenia: 10.10.2018

doc. RNDr. Damas Gruska, PhD.
garant študijného programu

študent

vedúci práce

Čestné prehlásenie

Čestne prehlasujem, že túto bakalársku prácu som vypracoval samostatne pod vedením vedúceho bakalárskej práce a s použitím uvedených zdrojov.

V Bratislave dňa 24.5.2015

.....

Pavol Freivolt

Pod'akovanie

Chcel by som sa pod'akovať môjmu vedúcemu práce Mgr. Petrovi Nátherovi, PhD. za jeho ochotu, trpezlivosť, rady a svojej rodine a kamarátom za morálnu podporu.

Abstrakt

Študujeme hru go z perspektívy komplexnej siete. Konštruujeme orientovaný graf za použitia vhodnej definície taktických ťahov a študujeme túto sieť na rôznych dátach hier profesionálov. Distribúcia ťahov sa riadi Zipfovým pravidlom a sieť je bezškálová, rovnako ako v prípade siete go. Štúdium vlastných hodnôt a vlastných vektorov matíc poukazujú na strategické situácie. Naša práca by mala prispieť k lepšiemu chápaniu šachu a strategickému mysleniu.

Abstract

We study the game of chess from a complex network perspective. We construct a directed network using a suitable definition of tactical moves, and study this network for different datasets of professional games. The distribution of moves follows Zipf's law and the network is scale free, same as the case of the go network. The study of eigenvalues and eigenvectors of matrices shows strategic situations. Our study should contribute to a better understanding of chess and strategic thinking.

1. Obsah

Úvod.....	2
1. Úvod do problematiky.....	3
2.1 Teória Grafov	3
2.2 Šach	4
2.3 Odborné štúdie vykonané predomnou	4
2.3.1 Od šachu po sieť	4
2.3.1 Od go po sieť	6
2. Cieľ práce	7
3. Použité technológie	8
4. Herná sieť šachu	9
5.1 Vstupné súbory	9
5.1.1 Tag pairs	9
5.1.2 Move text	9
5.1.3 Ukážka PGN súboru	9
5.2 Tvorba Grafu	10
5.2.1 Reprezentácia ťahov v sieti	10
5.3 Analýza	12
5.3.1 Najčastejšie ťahy	12
5.3.2 Najčastejšie sekvencie ťahov	16
5.3.3 Analýza siete	17
5. Záver	22
6. Zdroje	23

1. Úvod

So štúdiami komplexných sietí sa v posledných rokoch roztrhlo vrece. Táto zvedavosť je poháňaná najmä rozvojom v oblasti komunikácie a informatiky. Siete nám slúžia ako nástroj na hĺbkovú analýzu nielen v týchto oblastiach ale aj v mnohých iných. Čím ďalej, tým viac sa ukazuje spojenie reálneho sveta so svetom sietí. Siete sa dokonca dostali do takých odvetví, ako sociológia či jazyky.

Doskové hry sú jednou z najstarších aktivít ľudstva. Okrem krátenia času, slúžia tieto hry ako ukazovateľ aspektov ľudského myslenia a rozhodovania. Niektoré vlastnosti týchto hier je však ťažké pozorovať, ale to neznamená, že by sme sa o to nemali pokúšať. Nie je to tak dávno čo sme sa dostali na takú úroveň chápania týchto hier, že sa nám dokonca podarilo naučiť počítač ako v nich poraziť človeka. Jedna z týchto doskových hier, známa ako šach, je tejto práci analyzovaná a porovnávaná s doskovou hrou go. I keď obe hry majú iný princíp, iné pravidlá a iný počet možných ťahov, pokúsime sa v poukázať na podobnosti týchto hier.

V prvej kapitole si v krátkosti priblížime časti teórie grafov používané v tejto práci, povieme si niečo o šachu a budeme sa venovať aj prácam ktoré inšpirovali a napomohli vzniku práce. Následne vytýčime cieľ práce a spomenieme technológie, ktoré sme využili či už pri samotnom zostrojení našej siete alebo pri analýze. Nasleduje kapitola ktorá sa už venuje samotnému cieľu práce. Táto časť obsahuje opis narábania so vstupnými súbormi, myšlienkový postup pri tvorbe našej siete, od omylov k finálnej verzii, a samotné výsledky analýz tejto siete porovnané s výsledkami analýz siete go. Na záver zhrnieme čo sa udialo v našej práci a poskytneme čitateľovi zdroje, kde môže nájsť odpovede na prípadné otázky a nejasnosti.

2. Úvod do problematiky

2.1 Teória grafov [1]

Teória grafov je vedná oblasť zaoberajúca sa analýzou grafov. Graf je matematická štruktúra používaná na modelovanie relácií medzi objektami. Graf tvorený z vrcholov a hrán. Môže byť orientovaný alebo neorientovaný.

Vrchol [2]

Z pohľadu teórie grafu, sú vrcholy považované za beztvaré nedeliteľné objekty, hoci môžu mať štruktúru v závislosti od aplikácie, z ktorej graf vzniká. Sémantická sieť [3] je graf, v ktorom sú vrcholy konceptami alebo triedami objektov.

Hrana [4]

Hrana minimálne dva vrcholy. V prípade orientovaného grafu tvoria hrany usporiadaný pár, teda ak máme vrchol A a vrchol B potom hrana (A, B) je odlišná od hrany (B, A), kdežto v prípade neorientovaného grafu sú tieto hrany totožné.

Váha [5]

Číselná hodnota, priradená vrcholu alebo hrane grafu. Váha podgrafu je súčtom váh vrcholov alebo hrán v rámci tohto podgrafu.

Stupeň [6]

V teórii grafov, stupeň vrcholu je počet hrán doliehajúcich k danému vrcholu. Pri orientovanom grafe sa rozlišujú vstupné a výstupné hrany. Vstupné hrany sú také, ktoré sú orientované smerom k vrcholu, a výstupné sú tie, ktoré sú orientované smerom od vrcholu.

Klastrovací koeficient [7]

Klastrovací koeficient vyjadruje, do akej miery sa vrcholy v grafe majú tendenciu zoskupovať. Môže sa merať globálne alebo lokálne. Globálny klastrovací koeficient poskytuje celkovú informáciu o klastrí v sieti a lokálny udáva vstavanosť jednotlivých uzlov.

PageRank [8]

PageRank je algoritmus používaný vyhľadávačom Google na ohodnotenie webových stránok. Funguje tak, že spočíta počet odkazov na stránku a určí hrubý odhad toho aká je webová stránka dôležitá. Dá sa použiť aj v iných sieťach na odhadnutie dôležitosti vrcholu.

CheiRank [9]

CheiRank je PageRank s obrátenými hranami.

HITS [10]

Hyperlink-Inducted Topic Search alebo hubs and authorities je algoritmus na analýzu odkazov. Vychádza z konkrétneho nahliadnutia do tvorby webových stránok pri pôvodnom vytváraní internetu. Určité webové stránky, slúžili na kompiláciu katalógu informácií, ktoré viedli používateľov k iným stránkam.

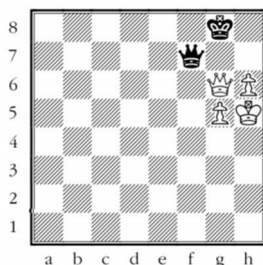
2.2 Šach

Šach je strategická dosková hra určená pre dvoch hráčov. I keď existujú aj iné variácie šachu, v tejto práci sa budeme zaoberať len tou bežnou. Hoci sa verí, že korene tejto hry siahajú až do siedmeho storočia, dodnes ju hrajú milióny ľudí po celom svete. Hracia plocha šachu obsahuje 64 políček (8x8). Stĺpce sú označené písmenami od a po h, riadky sú označené číslami od 1 po 8. Hrá sa s 32 figúrkami (16 bielych pre jedného hráča a 16 čiernych pre druhého). Existuje 6 druhov figúrok (kráľ, kráľovná, veža, strelec, kôň, pešiak) s rôznym spôsobom pohybu.

2.3 Odborné štúdie vykonané predom mnou

2.3.1 Od šachu po sieť

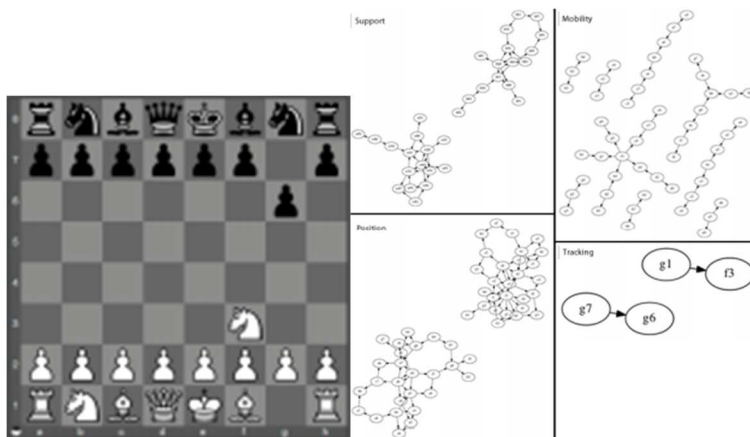
Na šach ako na komplexnú sieť sa ľudia snažia pozeráť už dlho. V práci Analysis of Network in Chess [11] nájdeme analýzu štyroch typov sietí. V mojej práci som síce nevyužil ani jednu z nižšie uvedených metód, ale pomohla mi pri tvorbe mojej metódy vyjadrenia ťahu v sieti.



Obrázok 1: Ukážka hernej pozície

(Obrázok 1) Za predpokladu, že figúrky sú vrcholy siete, za útočné hrany sa považujú tie, ktoré spájajú bielu kráľovnú s čiernou kráľovnou, bielu kráľovnú s čiernym kráľom a čiernu kráľovnú s bielou kráľovnou. Hrany, ktoré sa považujú za obranné sú tie, ktoré spájajú pešiakov, čierneho kráľa s čiernou kráľovnou, bielu kráľovnú s čiernym kráľom a opačne.

Koncepty týchto postupov boli použité na skonštruovanie štyroch rôznych sietí, ktoré modelujú rôzne aspekty hry.



Obrázok 2: Situácia na doske v ľavej časti obrázka je ukázaná vpravo na všetkých typoch nižšie spomínaných sietí.

Prvý typ je „support network“, ktorá reprezentuje vzťahy medzi figúrkami. Každý vrchol reprezentuje figúrku a má rovnakú farbu ako figúrka, ktorú reprezentuje. Vždy keď má figúrka možnosť vyhodit' figúrku inej farby, táto možnosť je znázornená pomocou hrany. Vždy keď má figúrka možnosť vyhodit' figúrku rovnakej farby (čo sa podľa pravidiel nedá), táto potencionálna možnosť je znázornená pomocou hrany. Hrany sú orientované a majú farbu podľa vrcholu, z ktorého vychádzajú. Každá hrana je buď útočná alebo obranná. Útočná je medzi vrcholmi rôznej farby, obranná medzi vrcholmi rovnakej farby.

Druhý typ je „mobility network“, ktorá reprezentuje ťah figúrok. Každý vrchol v takejto sieti je štvorec na hracej ploche, na ktorom je buď figúrka alebo je v dosahu figúrky. Každý vrchol, ktorý reprezentuje štvorec, na ktorom je figúrka, má rovnakú farbu ako táto figúrka. Ostaté vrcholy nemajú farbu. Hrany reprezentujú možnosť vyhodit' figúrku alebo pohnúť sa na daný štvorec. Hrany sú orientované a idú z vrcholu, ktorý reprezentuje štvorec, na ktorom je figúrka. Hrany medzi vrcholmi rovnakej farby nie sú. Každá hrana je útočná.

Tretí typ je „position network“, ktorá reprezentuje vzťahy medzi figúrkami a štvorcami, ktoré môžu obsadiť. Každý vrchol reprezentuje štvorec na hracej ploche. Každý vrchol, ktorý reprezentuje štvorec, na ktorom je figúrka, má rovnakú farbu ako táto figúrka. Ostaté vrcholy nemajú farbu. Táto sieť je zjednotenie „support network“ a „position network“.

Štvrtý typ je „tracking network“, ktorá reprezentuje históriu ťahov figúrok na hracej ploche. Každý štvorec na hracej ploche je reprezentovaný ako vrchol. Keď sa figúrka pohne z jedného štvorca na druhý, vznikne hrana medzi vrcholmi reprezentujúcimi tieto štvorce. Takáto hrana v sieti zostane až do konca hry.

2.3.2 Od go po sieť

Presne ako v prípade šachu, tak aj go sa snažia ľudia analyzovať. Ale zostrojiť sieť pre go je náročnejšie ako zostrojiť ju pre šach. V práci *The game of go as a complex network* [12], ktorou je moja práca inšpirovaná, takúto sieť zostrojili a zanalyzovali nižšie uvedeným spôsobom.

Go je dosková hra, ktorú hrajú dvaja hráči, čierny a biely. Doska sa skladá z 19 horizontálnych a 19 vertikálnych čiar. Hráči postupne pokladajú kamene na prázdne priesečníky čiar. Kameň, ktorý je úplne obkolesený kameňmi protivníka, musí byť odstránený. Keďže doska go je veľká, je ju potrebné rozdeliť na štyri 3x3 plakety. Plakety, ktoré sa navzájom premietajú pri akejkoľvek symetrii štvorca (rotácia alebo prevrátenie), sú považované za totožné a tiež ťah, ktorý čierny hráč hrá na danej plakete, je považovaný za ťah, ktorý biely hráč hrá na rovnakej plakete s vymenenými farbami. Takto vzniklo 1107 neekvivalentných plaket.

Dynamika hry je postavená na po sebe idúcich ťahoch. Preto na zostrojenie ich siete je potrebné spojiť dva ťahy iba ak išli po sebe v rovnakej oblasti na hracej doske. Sieť je zostrojená z 1107 vrcholov (ťahov) a tieto vrcholy sú poprepájané hranami s váhou, ktorá je určená počtom výskytov v databáze.

Na takto skonštruovanej sieti sú zanalyzované frekvencie ťahov, frekvencie sekvencií dvoch až siedmich po sebe idúcich ťahov a stupne vrcholov. Aby sa dalo sledovať ako sieť zachytáva aspekty stratégie hry, je použitá ohodnotená matica susedností (hrany sú ohodnotené podľa ich frekvencie v databáze). Kým PageRank a Hubs poukazujú na vlastnosti vrcholov závisiace od vstupných hrán, CheiRank a Authorities sú založené na výstupných hranách. Ďalej je analyzovaná závislosť vlastných hodnôt matice od počtu hier s ktorých bola táto matica vytvorená.

3. Cieľ práce

Cieľom tejto práce je zostrojenie siete vytvorenej zo šachových partií, analýza a vizualizácia rôznych aspektov tejto siete a porovnanie výsledkov, ku ktorým sme sa dopracovali, s výsledkami práce *The game of go as a complex network* [12].

4. Použité technológie

Python 3.7

Python je open source multi-paradigmaticý interpretačný jazyk schopný systémových volaní. Podporuje objektovo orientované, štruktúrované aj funkcionálne programovanie.

igraph

Igraph je open source knižnica na manipulovanie grafov a analýzu sietí. Je napísaná v jazyku C, ale existuje aj ako Python a R programový balík. Táto knižnica bola použitá pri implementácii našej siete.

numpy

Numpy je open source knižnica pre programovací jazyk Python, ktorá pridáva podporu pre obrovské multidimenzionálne polia a matice, a tiež kolekciu matematických funkcií. Túto knižnicu sme použili pri výpočte vlastných hodnôt matice a pri hľadaní sklonov funkcií.

python-chess

Python-chess je open source šachová knižnica pre programovací jazyk Python. Slúži na generovanie/validáciu ťahov a čítanie/zapisovanie PGN súborov. Knižnica bola v tejto práci použitá na čítanie z PGN súborov.

xlwt

Xlwt je open source knižnica na tvorbu tabuľkových súborov a za týmto účelom sme ju použili v tejto práci.

Gephi

Gephi je open source softvérový balíček na analýzu sietí a vizualizáciu. Pomocou Gephi sme vytvorili vizualizácie siete použité v práci (*Obrázok 4 a 6*).

gnuplot

Gnuplot je open source program pre generovanie 2D a 3D grafov a funkcií či dát. Výsledok vykresľuje na obrazovku v reálnom čase. Aj pomocou gnuplotu sme vytvorili vizualizácie použité v tejto práci (*Obrázok 9, 10 a Obrázok 12 - 20*).

5. Herná sieť šachu

5.1 Vstupné súbory

Predtým ako vytvoríme graf, potrebujeme dáta, z ktorých bude zostrojený. Šachové partie sa zvyknú zapisovať do súboru s formátom PGN (Portable Game Notation). Tento formát je navrhnutý tak aby sa ľahko modifikoval, či už zo strany človeka alebo zo strany počítača. Šachové ťahy sú reprezentované v algebrickej šachovej notácii. V PGN špecifikácii existujú dva formáty, import formát a export formát. Import formát môže byť aj ručne pripravený, ale export formát je striktná verzia, ktorú vytvorí počítač. PGN kód vždy začína s tag pairs (5.1.1) a za nimi nasleduje move text (5.1.2).

5.1.1 Tag pairs

Tag pairs začína zátvorkou [a končí zátvorkou]. V priestore medzi zátvorkami sa nachádza názov tagu a hodnota tagu. PGN dáta určené pre archívne účely sú povinné poskytnúť STR (Seven Tag Roster) a to Event (názov podujatia), Site (miesto konania podujatia), Date (dátum konania podujatia), Round (číslo kola), White (meno hráča, ktorý hrá za biele figúrky), Black (meno hráča, ktorý hrá za čierne figúrky) a Result (výsledok hry).

5.1.2 Move text

Move text popisuje ťahy, ktoré boli v hre vykonané. Toto zahŕňa ťahové číselné indikátory a SAN (Standard Algebraic Notation) [16]. SAN sa skladá zo skratky pre figúrku a z pozície, na ktorú bola figúrka posunutá.

5.1.3 Ukážka PGN súboru

```
[Round "1"]
[White "Edwardsen,R"]
[Black "Carlsen,Magnus"]
[Result "1/2-1/2"]
[WhiteElo "2055"]
[BlackElo ""]
[ECO "D12"]

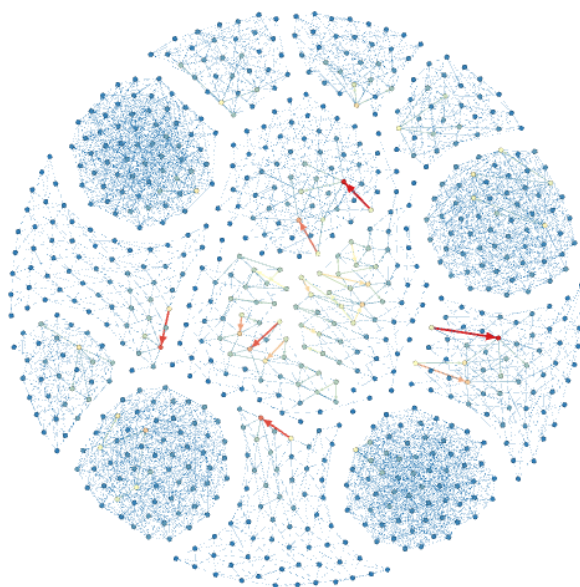
1.d4 Nf6 2.Nf3 d5 3.e3 Bf5 4.c4 c6 5.Nc3 e6 6.Bd3 Bxd3 7.Qxd3 Nbd7 8.b3 Bd6
9.O-O O-O 10.Bb2 Qe7 11.Rad1 Rad8 12.Rfe1 dxc4 13.bxc4 e5 14.dxe5 Nxe5 15.Nxe5 Bxe5
16.Qe2 Rxd1 17.Rxd1 Rd8 18.Rxd8+ Qxd8 19.Qd1 Qxd1+ 20.Nxd1 Bxb2 21.Nxb2 b5
```

Obrázok 3: Prvých sedem riadkov na ukážke zaberajú tag pairs (5.1.1). Dole na obrázku sa nachádza move text (5.1.2) písaný v SAN.

5.2 Tvorba Grafu

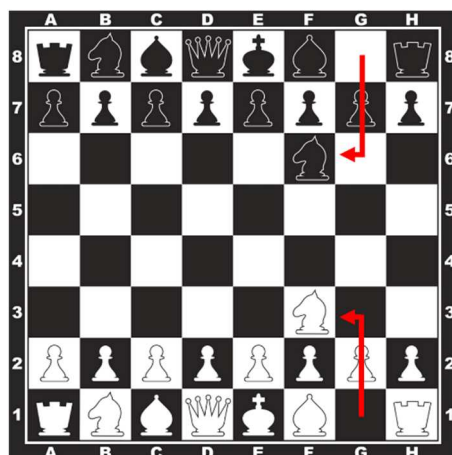
5.2.1 Reprezentácia ťahov v sieti

Testovali sme dva rôzne spôsoby vytvorenia grafu zo šachových partií. V oboch prípadoch bol graf orientovaný. Na začiatku sme zvolili spôsob, kde vrchol bol určený figúrkou, farbou figúrky a aktuálnou pozíciou figúrky. Ťah bol reprezentovaný dvojicou vrcholov spojenou ohodnotenou hranou. Ohodnotenie hrany bolo rovné počtu výskytov ťahu v databáze. Po prvom otestovaní sme prišli na to, že namiesto jedného grafu nám vzniklo štrnásť grafov (*Obrázok 4*) a to preto, lebo pre bieleho aj pre čierneho hráča existuje šesť typov figúrok, a keďže sa strelci rovnakej farby nikdy nemôžu hýbať po rovnakých pozíciách, môžeme ich brať za rôzne figúrky. Avšak tento počet grafov nám nevyhovuje, pretože chceme skúmať šach ako celok a nie z pohľadu jednotlivých figúrok.



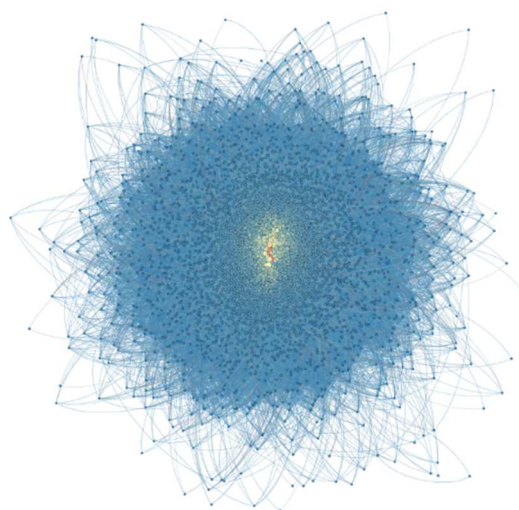
Obrázok 4: Grafické znázornenie prvej verzie grafu.

Museli sme teda nájsť takú implementáciu grafu, ktorá po načítaní dát vytvorí jeden graf. Rozhodli sme sa pre metódu, kde ťah nie je spojenie dvoch vrcholov ako v predchádzajúcom prípade, ale každý jeden ťah je reprezentovaný ako vrchol grafu a hrany spájajúce tieto vrcholy reprezentujú sekvenciu dvoch za sebou idúcich ťahov. Po následnom testovaní tejto štruktúry grafu sme zistili, že nám namiesto jedného grafu vzniknú dva, a to pre ťahy čiernych figúrok a ťahy bielych figúrok. Tento problém sme vyriešili tak, že sme ku každému ťahu bieleho hráča našli presne jeden ekvivalentný ťah čierneho hráča. Táto ekvivalencia bola určená zobrazením polovice šachovnice čierneho hráča do polovice šachovnice bieleho hráča podľa stredovej osi (*Obrázok 5*). Všetky ťahy čierneho hráča reprezentujeme v našom grafe ako ich ekvivalentné ťahy bieleho hráča.



Obrázok 5: Grafické znázornenie ekvivalentných ťahov bieleho a čierneho hráča.

Okrem figúrky, farby a pozície si vrchol grafu uchováva aj frekvenciu ťahu, ktorá je určená počtom výskytov ťahu v databáze. Hrany majú za váhu frekvenciu, ktorá je určená počtom výskytov sekvencie dvoch ťahov v databáze. Touto reprezentáciou ťahov sme konečne dospeli k jednému ucelenému grafu (Obrázok 6), na ktorom teraz môžeme vykonať analýzu šachových partií.



Obrázok 6: Grafické znázornenie grafu vytvoreného z cca. 25 000 hier profesionálnych hráčov od roku 1850 až po súčasnosť. Graf má cca. 3 800 vrcholov a cca. 650 000 hrán. Farba vrcholu je určená počtom hrán orientovaných do tohto vrcholu na farebnej škále od modrej po červenú, kde modrá reprezentuje nižší počet hrán orientovaných do vrcholu a červená vyšší. Farba hrany je určená jej váhou. Čím vyššia váha tým červenia hrana.

5.3 Analýza

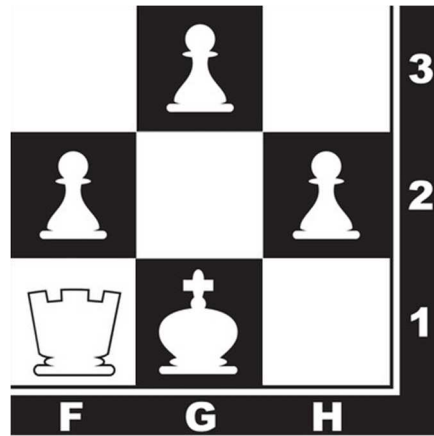
Všetky nami vyrobené analýzy v tejto práci sú vypracované na databázach pre troch najlepších hráčov Carlsen, Caruana, Ding a na databáze, na ktorú sa budeme referovať ako na databázu všetkých hráčov, kde sú hry stoviek hráčov vrátane Carlsena, Caruanu a Dinga. Farby funkcií v nasledujúcich grafoch pre šach sú: Carlsen - modrá, Caruana - červená, Ding – zelena, všetci hráči – žltá, ak nie je uvedené inak v popise pod obrázkom. Pre go: Honinbo - čierna, Meijia - červená, Judan – zelená, Kisei – modrá, amatérske turnaje - fialová, ak nie je uvedené inak v popise pod obrázkom. Logaritmus je dekadický, ak nie je uvedené inak v popise pod obrázkom. Všetky analýzy siete go, na ktoré sa v práci odvolávame, pochádzajú z práce The game of go as a complex network [12].

5.3.1 Najčastejšie ťahy

V šachu ide hlavne o ochranu kráľa a o ovládanie stredu šachovnice. Všetky ťahy v *Tabuľka 1* sú logickým odôvodnením tejto stratégie. Ťah O-O je perfektný ako na ochranu kráľa, tak aj na uvoľnenie veže a ak je tento ťah skombinovaný s ťahom Pg2-Pg3, vznikne veľmi silná obranná formácia (*Obrázok 7*), kde pešiak na f2 a pešiak na h2 kryje pešiaka na g3 a kráľ kryje pešiakov na f2 a h2. Pri ťahoch Pe2-Pe4, Pd2-Pd4 ide hlavne o ovládnutie stredu šachovnice a spolu s ťahmi Ng1-Nf3, Nb1-Nc3 sú použité v rôznych variáciách otvorení šachových partií ako napríklad Ruy Lopez, Italian game, Sicilian defense : Closed (*Obrázok 8*) a v mnohých ďalších.

Ph2-Ph3	18 006	0,95%
Pa2-Pa3	18 212	0,96%
Pg2-Pg3	20 956	1,11%
Pe2-Pe3	21 298	1,13%
Pe2-Pe4	23 317	1,23%
Pc2-Pc4	26 210	1,39%
Pd2-Pd4	32 494	1,72%
Nb1-Nc3	32 590	1,72%
O-O	38 249	2,02%
Ng1-Nf3	43 054	2,28%

Tabuľka 1: Najčastejšie ťahy z databázy všetkých hráčov



Obrázok 7

Ruy Lopez (Spanish game)



King's Indian defense



Italian game



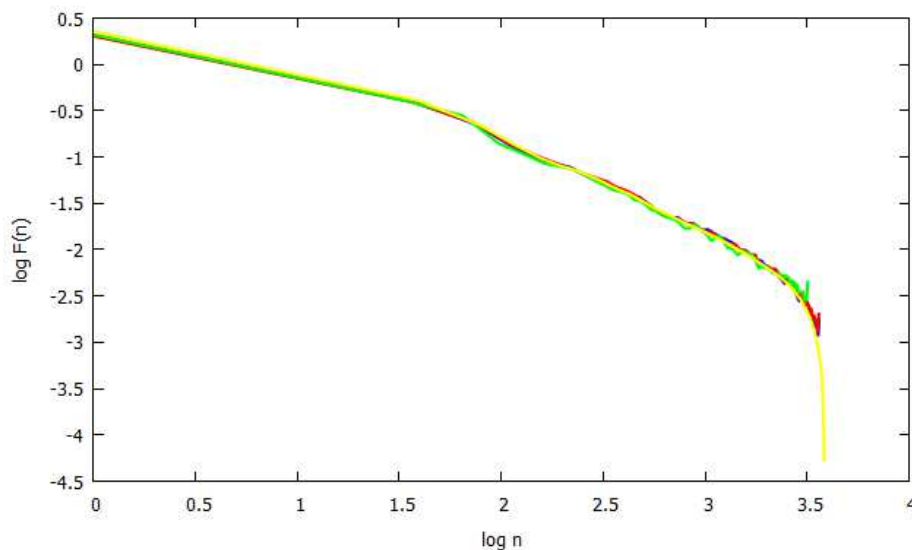
Sicilian defense: Closed

Obrázok 8 : Frekventované šachové otvorenia

Ding			Carlsen			Caruana		
Ph2-Ph3	396	0,88%	Pa2-Pa3	994	0,87%	Rf1-Re1	809	0,81%
Bf1-Bg2	420	0,94%	Pe2-Pe3	1 027	0,90%	Pa2-Pa3	841	0,85%
Pe2-Pe4	435	0,97%	Ph2-Ph3	1 046	0,92%	Ph2-Ph3	991	1,00%
Pe2-Pe3	508	1,13%	Pg2-Pg3	1 190	1,05%	Pc2-Pc4	1 018	1,03%
Pg2-Pg3	572	1,28%	Pe2-Pe4	1 355	1,19%	Pg2-Pg3	1 020	1,03%
Pc2-Pc4	680	1,52%	Pc2-Pc4	1 373	1,21%	Pe2-Pe4	1 243	1,25%
Pd2-Pd4	717	1,60%	Pd2-Pd4	1 619	1,42%	Pd2-Pd4	1 379	1,39%
Nb1-Nc3	726	1,62%	Nb1-Nc3	1 728	1,52%	Nb1-Nc3	1 421	1,43%
O-O	851	1,90%	O-O	2 086	1,84%	O-O	1 732	1,74%
Ng1-Nf3	936	2,09%	Ng1-Nf3	2 347	2,07%	Ng1-Nf3	2 000	2,01%

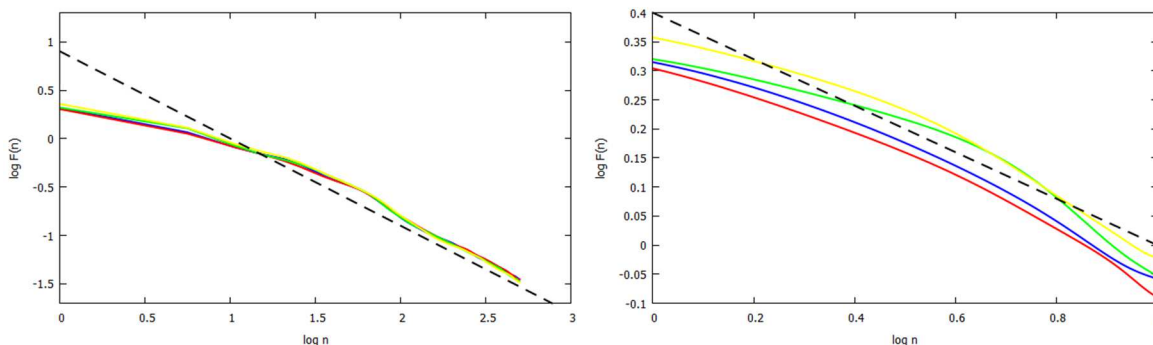
Tabuľka 2: Najčastejšie ťahy z databázy hráčov Carlsen, Caruana a Ding.

Keď sa pozrieme na *Tabuľka 2*, môžeme si všimnúť, že frekvencia najčastejších ťahov pre hráčov Carlsen, Caruana, Ding je nižšia ako na *Tabuľka 1*. Veta č. 1: To indikuje, že títo hráči hrajú menej až stredne frekventované ťahy častejšie ako je to v prípade priemerného hráča z našej databázy všetkých hráčov. Ak porovnávame týchto hráčov medzi sebou vidíme, že ich štyri najčastejšie ťahy sú totožné. Ťah Bf1-Bg2 u Dinga môže naznačovať, že často hrá otvorenia ako King's Indian defense (*Obrázok 8*) alebo Owen's Defence. Ťah Rf1-Re1 u Caruanu naznačuje, že po malej rošáde sa snaží uvoľniť si vežu a kryť ňou figúrky v strede šachovnice.



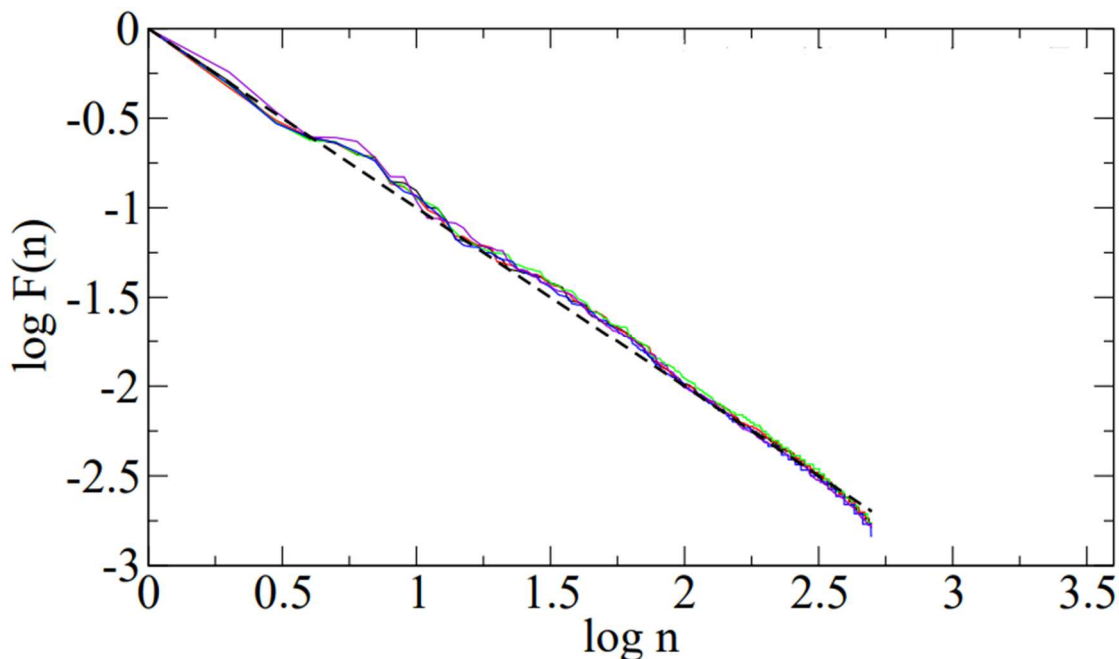
Obrázok 9: (Šach) Normalizovaná integrovaná distribúcia frekvencií ťahov $F(n)$. Ťahy z databázy všetkých hráčov sú zoradené od najfrekventovanejšieho po najmenej frekventovaný a ťahy z databázy pre hráčov Carlsen, Caruana a Ding sú zoradené podľa ťahov z databázy všetkých hráčov. Logaritmus je prirodzený.

Z *Obrázok 9* vyplýva, že Carlsen, Caruana a Ding sa pri prvých približne 50-tich najčastejších ťahov príliš nelíšia od priemerného hráča z databázy všetkých hráčov, ale potom je odchýlka čoraz väčšia. Prekvapujúce je zistenie, že táto trojica hráčov využíva menej frekventované ťahy častejšie ako priemerný hráč z databázy všetkých hráčov. To potvrdzuje našu hypotézu z vety č. 1.



Obrázok 10: Vľavo: Normalizovaná integrovaná distribúcia frekvencií 500 najčastejších ťahov $F(n)$. Sklony jednotlivých funkcií sú: Carlsen -0.849 , Caruana -0.839 , Ding -0.866 a všetci hráči -0.884 . Prerušovaná čiara je $y = -0.9x + 0.9$. Vpravo: Normalizovaná integrovaná distribúcia frekvencií 10 najčastejších ťahov $F(n)$. Sklony jednotlivých funkcií sú: Carlsen -0.402 , Caruana -0.401 , Ding -0.391 a všetci hráči -0.408 . Prerušovaná čiara je $y = -0.4x + 0.4$.

Integrovaná distribúcia päťsto ťahov je podobá pre všetky naše databázy. Riadi sa Zipfovým pravidlom s exponentom ≈ 0.9 . Po vytvorení *Obrázok 10 Vľavo* sme si všimli, že prvých 10 ťahov sa neriadi týmto exponentom. Rozhodli sme sa ich merať zvlášť (*Obrázok 10 Vpravo*). Exponent pri týchto desiatich ťahoch bol ≈ 0.4 . Preskúmali sme aj exponent 500 najčastejších ťahov, nepočítajúc prvých 10 a vyšlo nám číslo ≈ 0.93 . Toto číslo sa blíži exponentu 1, ktorý môžeme sledovať, keď zoradíme mestá v USA od najväčšieho po najmenšie podľa počtu obyvateľov [13]. Podobný exponent bol nameraný pre frekvencie ťahov go, kde tento exponent nadobúda hodnotu ≈ 1.06 . Keď porovnáваме *Obrázok 10* s *Obrázok 11*, vidíme, že v prípade go, odchýlky prvých 10 ťahov od funkcie $y = -x$ sú výrazne menšie ako odchýlky prvých 10 ťahov od funkcie $y = -0.9x + 0.9$ v prípade šachu.



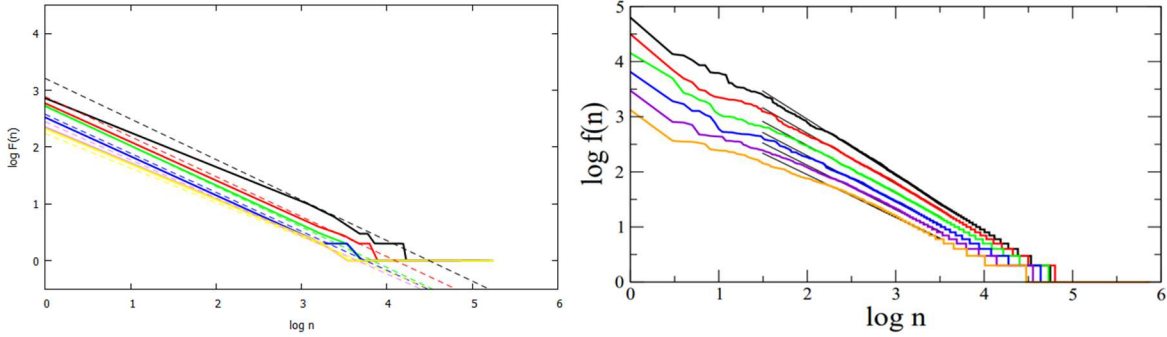
Obrázok 11: (Go) Normalizovaná integrovaná frekvencia distribúcií ťahov $F(n)$. Sklony funkcií sú: Honinbo -1.058, Meijia -1.056, Judan -1.065, Kisei -1.067 a amatérske turnaje -1.081. Prerušovaná čiara je $y = -x$.

5.3.2 Najčastejšie sekvencie ťahov

Pe2-Pe4 --> pe7-pe5
Pe2-Pe4 --> pe7-pe5 --> Ng1-Nf3
Pe2-Pe4 --> pe7-pe5 --> Ng1-Nf3 --> nb8-nc6
Pe2-Pe4 --> pe7-pe5 --> Ng1-Nf3 --> nb8-nc6 --> Bf1-Bb5
Pe2-Pe4 --> pe7-pe5 --> Ng1-Nf3 --> nb8-nc6 --> Bf1-Bb5 --> pa7-pa6
Pe2-Pe4 --> pe7-pe5 --> Ng1-Nf3 --> nb8-nc6 --> Bf1-Bb5 --> pa7-pa6 --> Bb5-Ba4

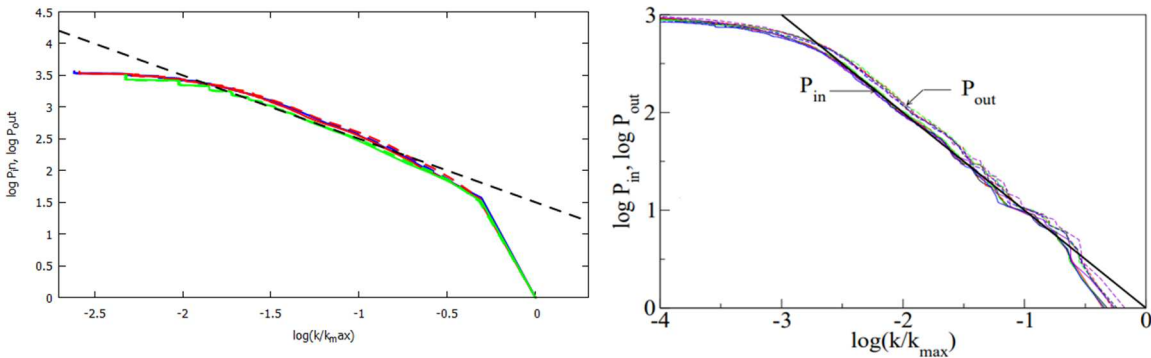
Tabuľka 3: Zhora nadol, najčastejšie vyskytujúca sa sekvencia 2 – 7 ťahov.

Z časti, ktorá ukazuje najčastejšiu sekvenciu šiestich ťahov môžeme vidieť, že ťahy bieleho hráča sú zhodné so sekvenciou ťahov prislúchajúcej k Ruy Lopez otvoreniu (Obrázok 8), taktiež známom pod menom Spanish game. Toto otvorenie je najviac frekventované a za ním nasledujú Italian game a Sicilian game (Obrázok 8) [14]. Ak porovnáme Obrázok 12 Vľavo s Obrázok 12 Vpravo môžeme vidieť, že počet najčastejších sekvencií v go je vyšší ako v prípade šachu. V go exponent siaha od ≈ 0.8 po ≈ 1 , kdežto v šachu je tento rozsah menší. Od ≈ 0.6 po ≈ 0.7 .



Obrázok 12: Integrovaná distribúcia frekvencií sekvencií ťahov $F(n)$ pre sekvenciu dvoch (čierna), troch (červená), štyroch (zelená), piatich (modrá), šiestich (fialová), siedmich (žltá) po sebe idúcich ťahov. *Vľavo:* (Šach) Sklony funkcií sú: čierna -0.713, červená -0.704, zelená -0.713, modrá -0.686, fialová -0.669 a žltá -0.607. *Vpravo:* (Go) Sklony funkcií sú: čierna -1.01, červená -0.91, zelená -0.86, modrá -0.83, fialová -0.81 a žltá -0.77.

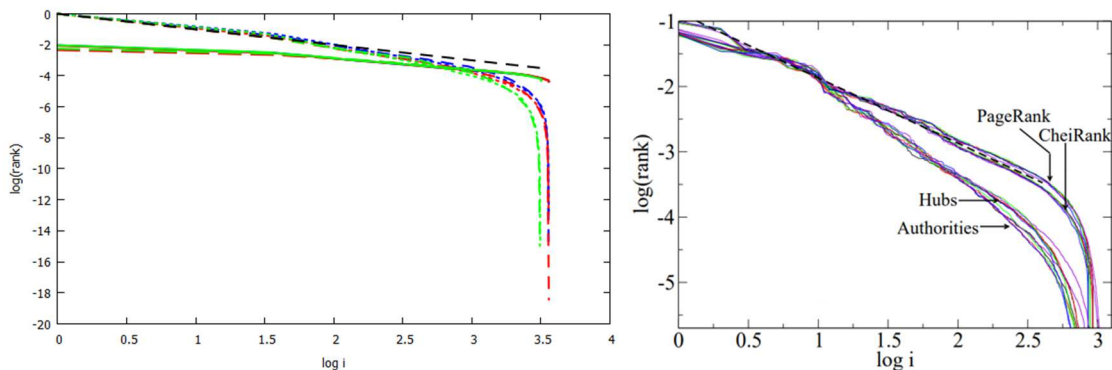
5.3.3 Analýza siete



Obrázok 13: Os y zachytáva počet vrcholov s viac než k vstupných/výstupných hrán a os x normalizovaný počet vstupných/výstupných hrán. Hodnota k je číslo ktoré bolo priradené vrcholu. Vrchol s najvyšším počtom vstupných/výstupných hrán má pridelené číslo 1 a vrchol s najnižším počtom k_{max}). *Vľavo:* (Šach) Normalizovaná integrovaná distribúcia vstupných P_{in} (spojené čiary) a výstupných P_{out} (prerušované čiary) hrán. Prerušovaná čiara predstavuje $y = -x + 1.5$. *Vpravo:* (Go) Normalizovaná integrovaná distribúcia vstupných P_{in} (spojené čiary) a výstupných P_{out} (prerušované čiary) hrán. Čierna spojená čiara predstavuje $y = -x$.

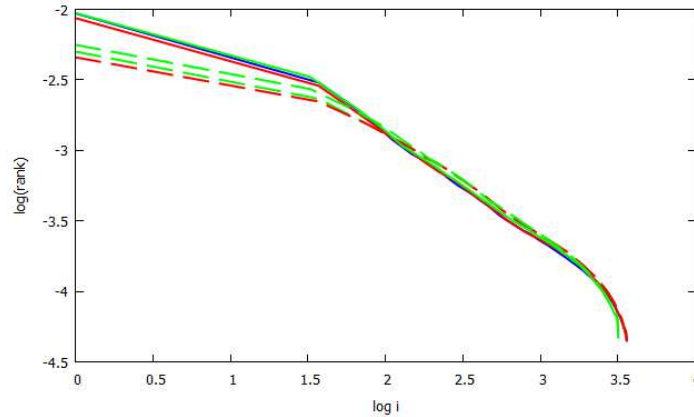
Integrované distribúcie oboch funkcií z *Obrázok 13* sú blízko k power-law $1/k^s$ s exponentom $s = 1$. Výsledky exponentu s sú nezávislé od použitej databázy. V oboch prípadoch siete zdieľajú vlastnosti bezškálovej siete [15]. Výsledky poukazujú na symetriu medzi ťahmi, ktoré často nasledujú iné ťahy a ťahmi, ktoré majú veľa možných následných ťahov. To nasvedčuje tomu, že mnoho krátkych sekvencií ťahov môže byť hraných v rôznom poradí v rámci rôznych kontextov.

Rovnako ako v prípade siete go, aj my vytvoríme ohodnotenú maticu susedností. Použijeme Google maticu G , definovanú ako $G_{ij} = \alpha S_{ij} + (1 - \alpha) e e^T / N$, kde $e = (1, \dots, 1)$, N je počet vrcholov, $0 < \alpha \leq 1$, S je ohodnotená matica susedností, kde každý stĺpec so súčtom 0 je nahradený stĺpcom obsahujúcim samé 1, a súčet každého stĺpca je normalizovaný na 1. Ale keďže v prípade siete go použili G , kde $\alpha = 1$, použijeme ju aj my. Teda vyššie uvedený vzorec môžeme upraviť na $G_{ij} = S_{ij}$.

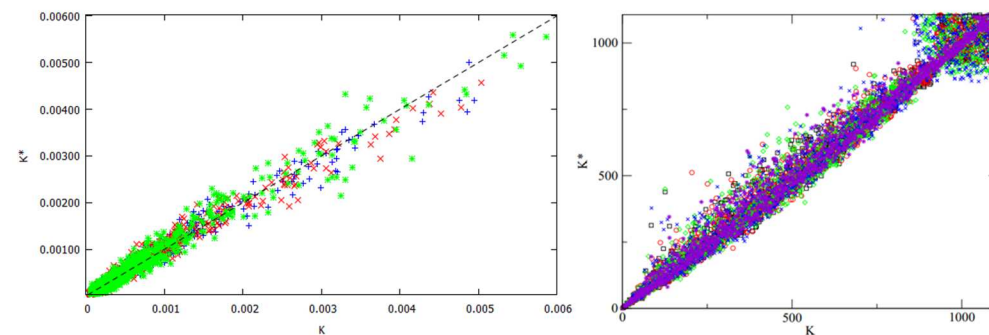


Obrázok 14: Hodnotiace vektory pre maticu G . Prerušovaná čiara zobrazuje funkciu $y = -x$. *Vľavo:* (Šach) Neprerušované čiary znázorňujú PageRank, prerušované dlhé čiarkované čiary CheiRank, prerušované krátke čiarkované čiary znázorňujú Hubs a prerušované bodkované čiary znázorňujú Authorities. *Vpravo:* (Go) Zhora nadol: PageRank, CheiRank, Hubs, Authorities.

Na *Obrázok 14* sme ukázali ohodnotenú maticu susedností pre siete vytvorené z našich databáz. V prípade go tieto sa tieto vektory riadia algebraickým pravidlom so sklonmi: PageRank, CheiRank ≈ -1 a Hubs, Authorities ≈ -1.5 . V prípade šachu sú sklony pre PageRank ≈ -0.7 , Hubs a Authorities ≈ -1.5 . Výsledky sklonu Hubs a Authorities pre sieť šachu sú totožné s výsledkami týchto ohodnotených vektorov pre sieť go. Pri PageRanku je sklon v prípade siete šachu nižší ako v prípade siete go, ale to môže byť spôsobené inou konfiguráciou PageRanku, nakoľko v práci The game of go as a complex network [12] nie je uvedená presná konfigurácia. Na *Obrázok 14* v oboch grafoch a na *Obrázok 15* si môžeme všimnúť významnú symetriu medzi ohodnotenými vektormi založenými na vstupných hranách s tými založenými na výstupných hranách.

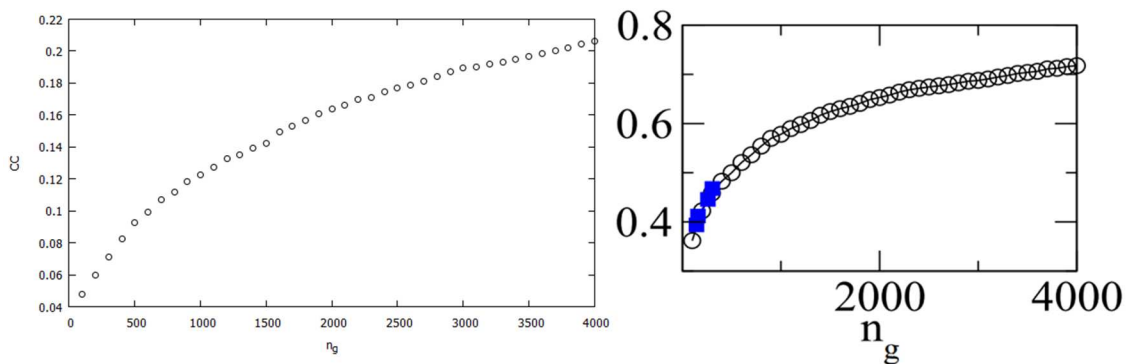


Obrázok 15: Príbliženie na PageRank a CheiRank z Obrázok 14



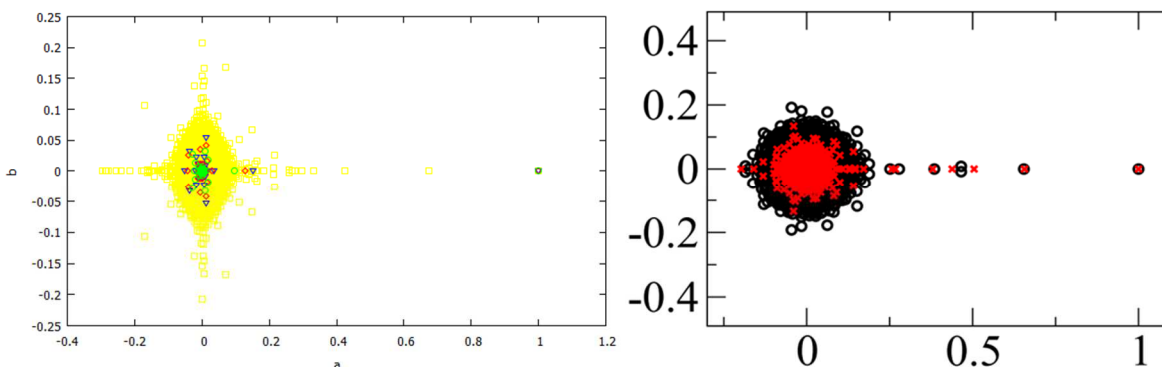
Obrázok 16: K^* vs K , kde K je ohodnotenie vrcholu zoradené podľa hodnoty PageRanku. Vľavo: (Šach) Prerušovaná čiara má predpis $y = x$. Vpravo: (Go).

Na to aby sme lepšie ukázali symetriu, postavili sme voči sebe PageRank s CheiRankom (Obrázok 16). Zistili sme, že existuje silná korelácia medzi týmito rankami založenými striktne na vstupných a výstupných hodnotách. čo naznačuje, že v šachu rovnako ako v go je silná korelácia medzi ťahmi, ktoré otvárajú veľa možností nových ťahov a ťahmi, ktoré môžu nasledovať veľa ďalších ťahov.



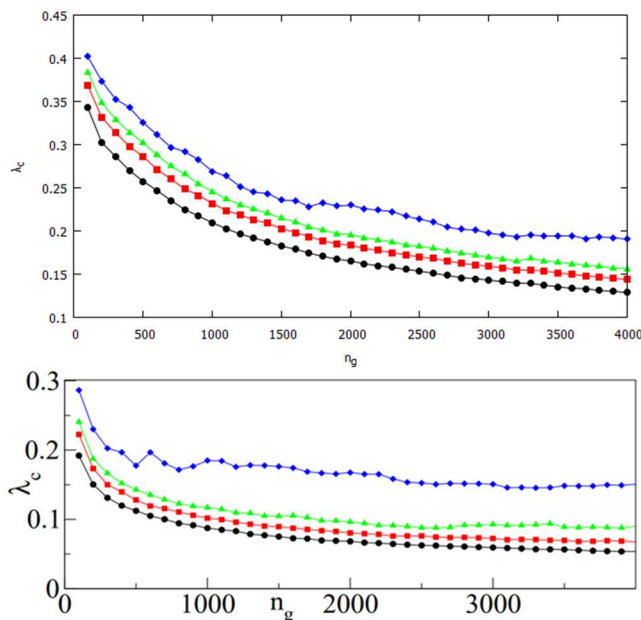
Obrázok 17: Klastrovací koeficient (CC) vs počet hier (n_g). Vľavo: (Šach) Na zostrojenie tohto grafu bola použitá databáza všetkých hráčov. Vpravo: (Go) Modré štvorce reprezentujú profesionálne turnaje a biele krúžky amatérske turnaje.

Podobne ako v sieti go aj v sieti šachu veľkosť klastrovacieho koeficientu závisí od počtu hier, ale v prípade go je tvorba klastrov výraznejšia (Obrázok 17).

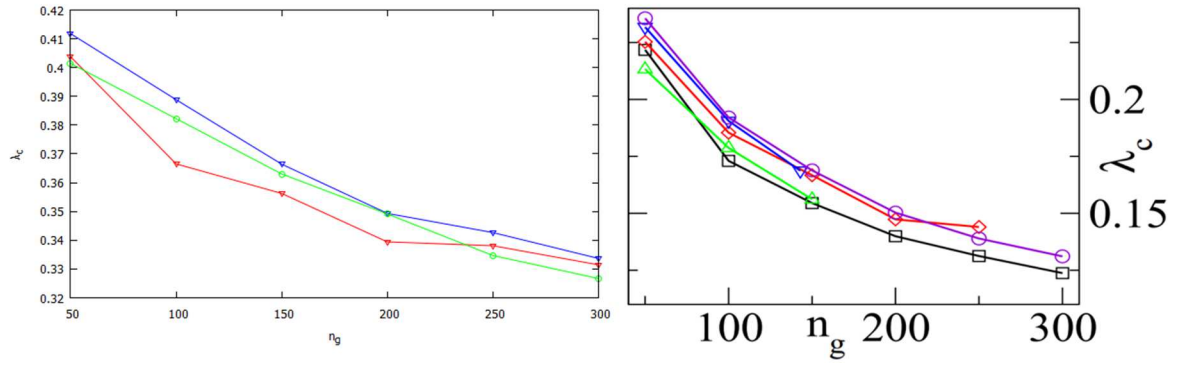


Obrázok 18: Vlastné hodnoty v komplexnej rovine pre maticu G . Os x predstavuje reálnu časť vlastnej hodnoty a os y imaginárnu. Vľavo: (Šach). Vpravo: (Go).

Rovnako ako v sieti go tak aj v sieti šachu je obrovská medzera medzi prvou vlastnou hodnotou $\lambda = 1$ a zhlukom vlastných hodnôt. Toto naznačuje, že sieť je veľmi prepojená s malými izolovanými komunitami. Tieto komunity sú tvorené sekvenciami taktických ťahov prednostne hraných spolu. Tieto komunity majú spoločné strategické vlastnosti. Obrázok 19 a Obrázok 20 ukazujú, že rádius zhluku vlastných hodnôt sa mení s množstvom hier. Toto indikuje, že čím viac hier sa berie do úvahy, tým viac zriedkavé hrany sa objavujú, čo rozbíja slabo zoskupené komunity.



Obrázok 19: λ_c taká, že zvrchu nadol 99%, 95%, 90%, 80% vlastných hodnôt λ , kde $|\lambda| < \lambda_c$. Hore: (Šach) Dáta čerpané z databázy všetkých hráčov. Dole: (Go) Dáta čerpané z databázy amatérskych hier.



Obrázok 20: λ_c 80% vlastných hodnôt λ , kde $|\lambda| < \lambda_c$. Vľavo: (Šach). Vpravo: (Go).

6. Záver

V tejto práci sme sa venovali hre šach, jednej z najstarších a najzložitejších doskových hier, z pohľadu komplexných sietí. Definovali sme správnu kategorizáciu ťahov a pomocou tejto kategorizácie sme vytvorili siete z hier stoviek profesionálov s bližším zameraním na najlepších troch Carlsen, Caruana a Ding. Ukázali sme prítomnosť Zipfovho pravidla vyskytujúceho sa v mnohých aspektoch takto vytvorených sietí. Medzi najviac frekvencovanými ťahmi sa vyskytujú najmä úvodné ťahy šachových partíí a šachové otvorenia. Poukázali sme na symetriu medzi vstupnými a výstupnými hranami, ktorá sa odzrkadľuje v symetrii ohodnotených vektorov. Pozorovali sme prítomnosť veľa malých izolovaných komúní, ktoré zdieľajú rovnaké strategické prvky. Zistili sme, že napriek rôznym herným pravidlám a konceptu, má nami skonštruovaná herná sieť šachu veľa spoločného s hernou sieťou go než by sa na prvý pohľad mohlo zdať. Túto hypotézu podporujú naše analýzy z kapitoly 5, ktoré sme porovnali s analýzami práce *The game of go as a complex network* [12]. Podobnosti sme našli v každom skúmanom aspekte siete.

Myslíme si, že naša práca napomôže lepšiemu chápaniu a modelovaniu hier šachu a zvýrazní podobnosti medzi strategickými doskovými hrami, čo môže inšpirovať ďalšie výskumy, ktoré sa budú uberať týmto smerom. Napríklad bolo by zaujímavé vidieť nami vypracované analýzy na ináč skonštruovanej sieti šachu s inou kategorizáciou ťahov alebo sa zamerať len na samotné pozície hracej dosky.

7. Zdroje

- [1] https://en.wikipedia.org/wiki/Graph_theory
- [2] [https://en.wikipedia.org/wiki/Vertex_\(graph_theory\)](https://en.wikipedia.org/wiki/Vertex_(graph_theory))
- [3] https://en.wikipedia.org/wiki/Semantic_network
- [4] https://en.wikipedia.org/wiki/Glossary_of_graph_theory_terms#edge
https://en.wiktionary.org/wiki/directed_graph
- [5] https://en.wikipedia.org/wiki/Glossary_of_graph_theory_terms#weighted_graph
- [6] [https://en.wikipedia.org/wiki/Degree_\(graph_theory\)](https://en.wikipedia.org/wiki/Degree_(graph_theory))
- [7] https://en.wikipedia.org/wiki/Clustering_coefficient
- [8] <https://en.wikipedia.org/wiki/PageRank>
- [9] <https://en.wikipedia.org/wiki/ChenRank>
- [10] https://en.wikipedia.org/wiki/HITS_algorithm
- [11] Derek Faaren, Daniel Templeton and Meiji Wang Team 23 Stanford University, *Analysis of Network in Chess*
- [12] B. Georgeot and O. Giraud, *The game of go as a complex network* (2012)
- [13] Xavier Gabaix , *The Quarterly Journal of Economics*, part Zipf`s law for Cities
- [14] <https://www.thesprucecrafts.com/most-common-chess-openings-611517>
- [15] https://en.wikipedia.org/wiki/Scale-free_network
- [16] [https://en.wikipedia.org/wiki/Algebraic_notation_\(chess\)](https://en.wikipedia.org/wiki/Algebraic_notation_(chess))