The network of go game

Vivek Kandiah, Bertrand Georgeot
Laboratoire de Physique Théorique, IRSAMC
CNRS/ University Paul Sabatier, Toulouse
Olivier Giraud
LPTMS CNRS/University Paris-Sud

Supported by EC FET Open project NADINE
CNRS and Région Midi-Pyrénées
B.G. and O. Giraud, Europhysics Letters 97 68002 (2012)
Vivek Kandiah, B.G. and O. Giraud, in preparation

Networks

Recent field: study of complex networks
->Tools and models have been created
->Many networks are scale-free with power-law distribution of links
->Difference between directed and non directed networks
->Important examples from recent technological developments: internet, World Wide Web, social networks...
->Can be applied also to less recent objects
In particular, study of human behavior: languages, friendships…

Networks for games

->Network theory never applied to games
->Games represent a privileged approach to human decision-making
->Can be very difficult to modelize or simulate
->While Deep Blue famously beat the world chess champion Kasparov in 1997,
no computer program has beaten a very good go player
 even in recent times.

Rules of go

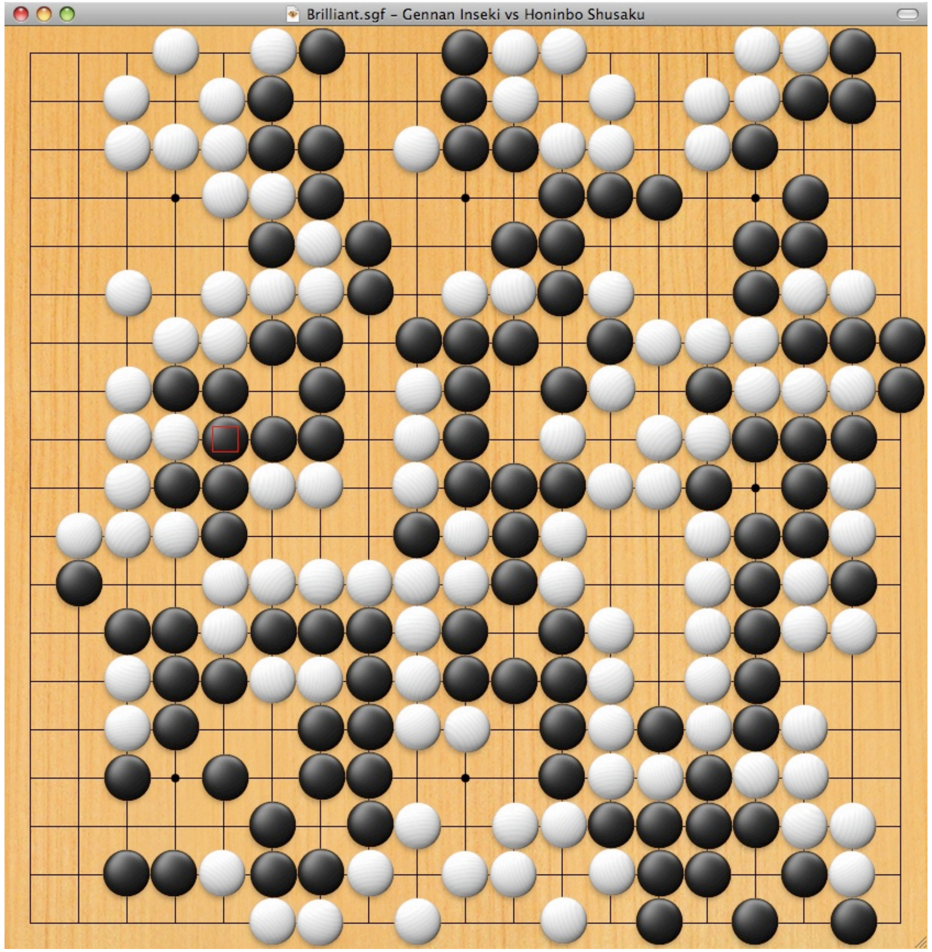
->White and black stones alternatively put at intersections of 19 x19 lines

->Stones without liberties are removed

->Handicap stones can be placed

->Aim of the game: construct protected territories

->total number of legal positions 10^{171}, compared to 10^{50} for chess



Databases

->We use databases of expert and amateur games in order to construct networks from the different sequences of moves, and study the properties of these networks

->Databases available at http://www.u-go.net/

->Whole available record, from 1941 onwards, of the most important historical professional Japanese go tournaments: Kisei (143 games), Meijin (259 games), Honinbo (305 games), Judan (158 games)

->First stage: to increase statistics and compare with professional tournaments, 4000 amateur games also used.

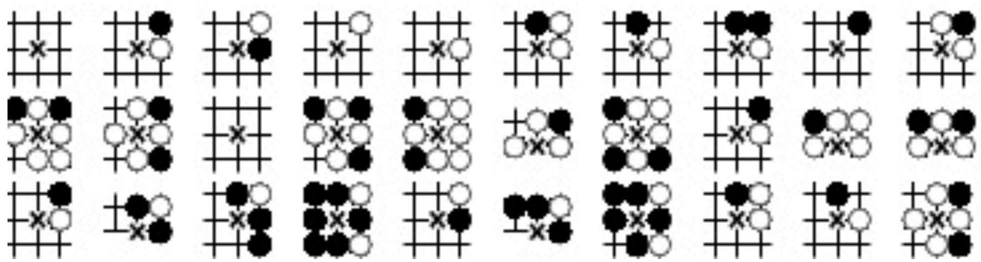
->Second stage: the whole database of 135 000 amateur games was used.

->Level of players is known

Vertices of the network I

->"plaquette" : square of 3 x3 intersections

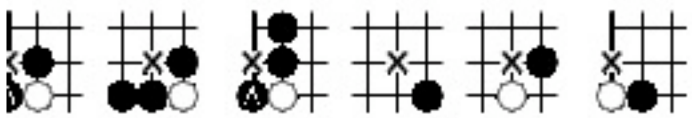
->We identify plaquettes related by symmetry
->We identify plaquettes with colors swapped
->1107 nonequivalent plaquettes with empty centers
->vertices of our network



Vertices of the networks II

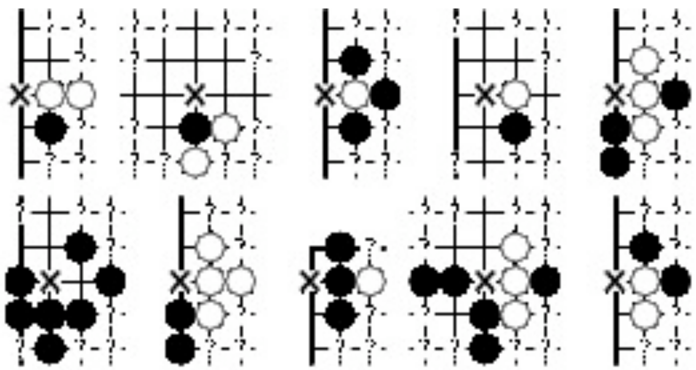
->"plaquette" : square of 3 x3 intersections + atari status of nearest-neighbors

->We still identify plaquettes related by symmetry
->2051 legal nonequivalent plaquettes with empty centers



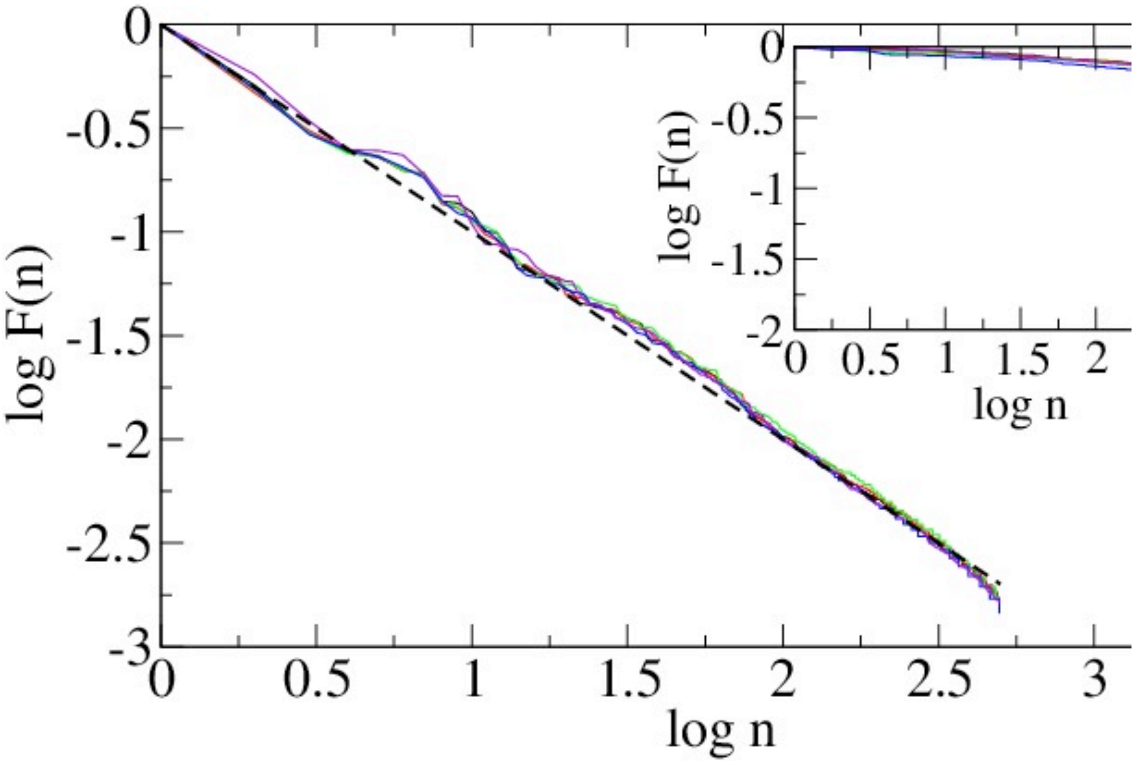
Vertices of the networks III

->"plaquette" : diamond of 3 x3 +4 intersections
->We still identify plaquettes related by symmetry
->193995 nonequivalent plaquettes with empty centers



Zipf's law

->Zipf's law: empirical law observed in many natural distributions (word frequency, city sizes...)
->If items are ranked according to their frequency, predicts a power-law decay of the frequency vs the rank.
->integrated distribution of 1107 moves clearly follows a Zipf's law,
with exponent 1.06

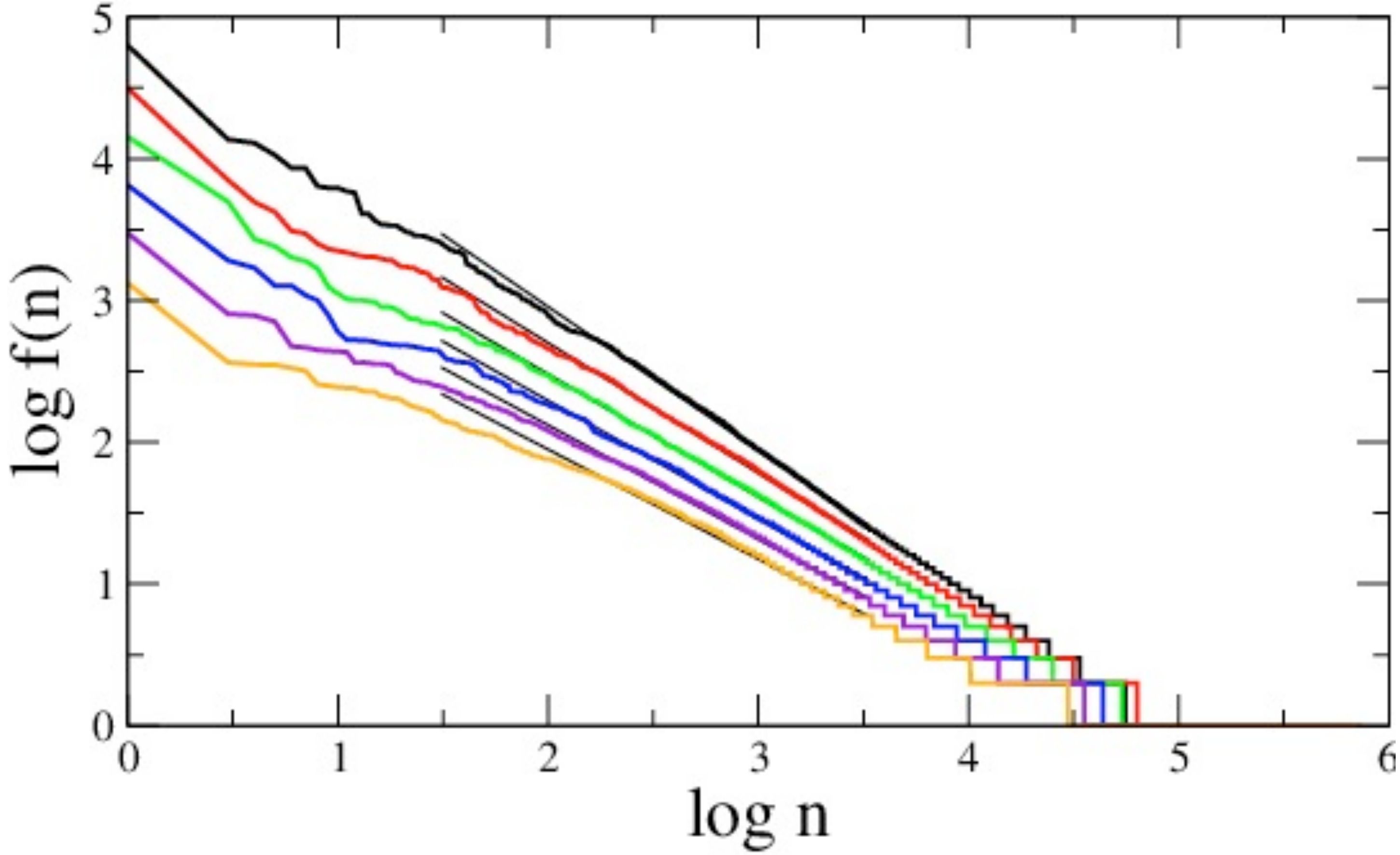


Normalized integrated frequency distribution of 1107 moves. Thick dashed line is y=-x.

Links of the network

->we connect vertices corresponding to moves a and b if b follows a in a game at a distance < d.
->Each choice of d defines a different network. The choice of d determines the distance beyond which two moves are considered nonrelated.
->Sequences of moves follow Zipf's law (cf languages) Exponent decreases as longer sequences reflect individual strategies
->move sequences are well hierarchized by d=5
->amateur database departs from all professional ones, playing more often at shorter distances

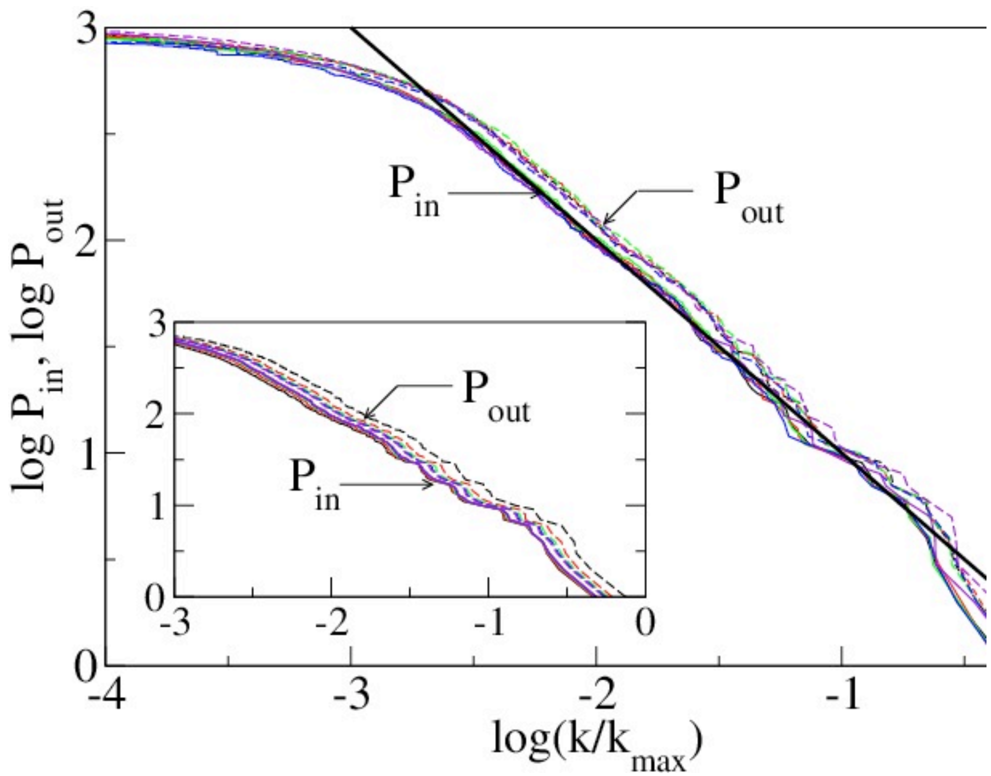
Links of the network



Integrated frequency distribution of sequences of moves f(n) for (from top to bottom) two to seven successive moves plotted against the ranks of the moves.

Links distribution

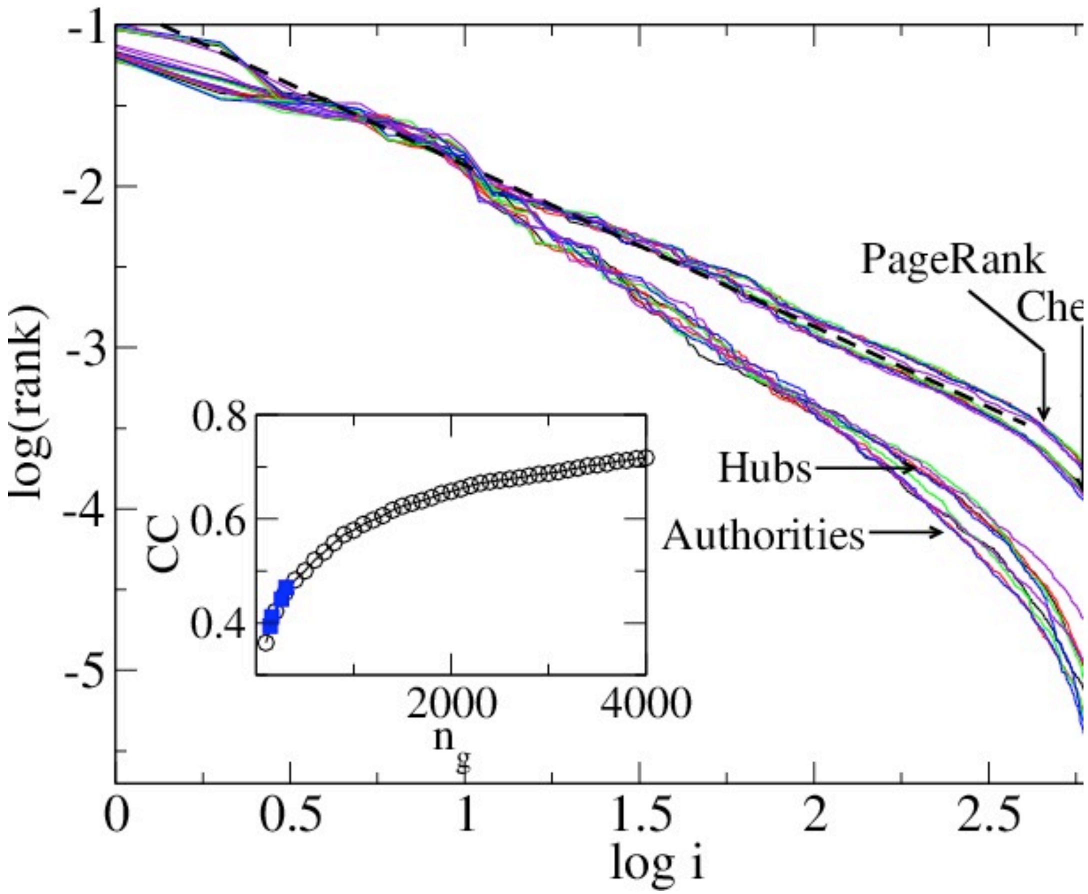
->Tails of link distributions very close to power-law with exponent 1.0 for the integrated distribution.
->The results are stable with respect to the database considered.
->network displays the scale-free property
->symmetry between ingoing and outgoing links is a peculiarity of this network



Normalized integrated distribution of links for d=5
Thick solid line is y=-x.
Inset:different values of d

Ranking vectors

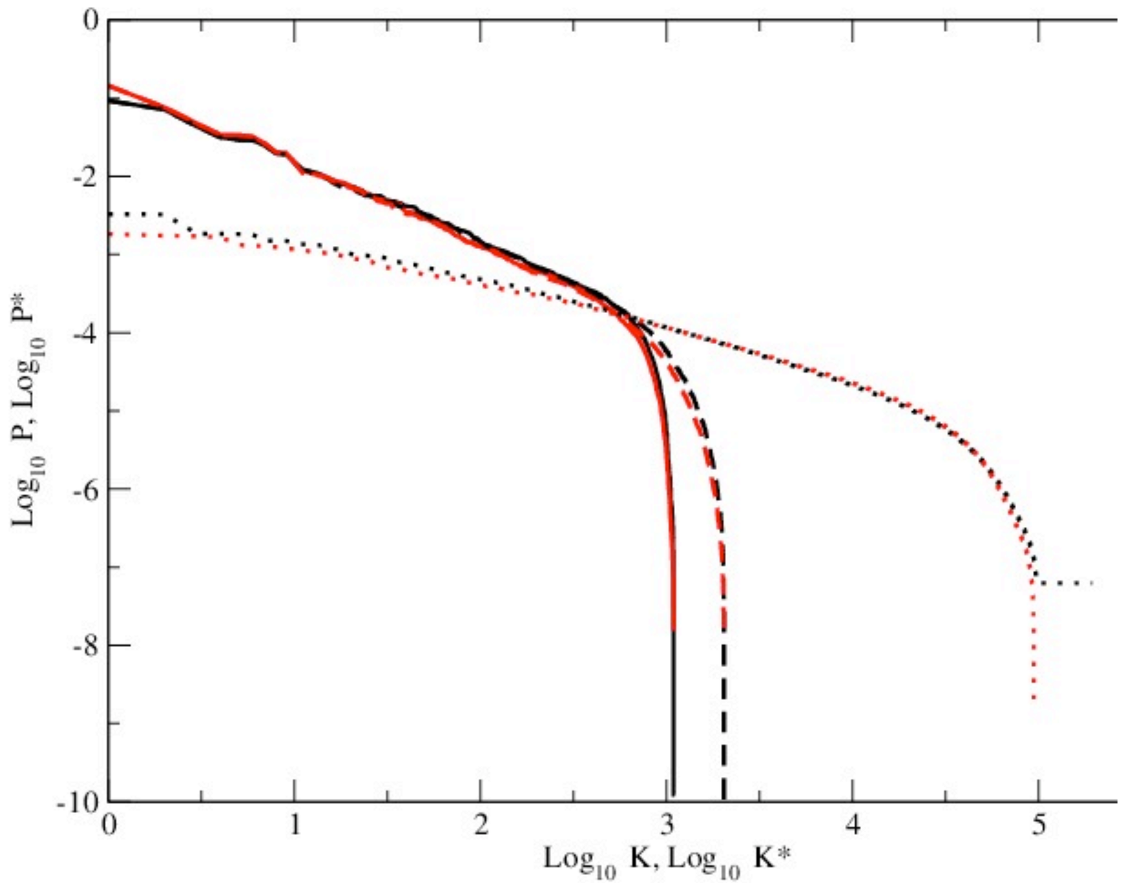
->PageRank: ingoing links
->CheiRank: outgoing links
->HITS algorithm:Authorities (ingoing links) and Hubs (outgoing links)
->Ranking vectors follow an algebraic law
->Symmetry between distributions of ranking vectors based on ingoing links and outgoing links.



->Clustering coefficient detects local connected clusters. Here depends on the number of games n_g included (see inset);for large n_g asymptotic value larger than 0.7 (WWW 0.11); CC = 0.7 with atari, CC=0.05 for diamond

Ranking vectors: other networks

->Still symmetry between distributions of ranking vectors based on ingoing links and outgoing links.
->Power law different for the largest network



->Ranking vectors of G. Black is PageRank, Red is CheiRank, Plain line: size 1107, dashed line: size 2051, dotted line: size 193995.

Ranking vectors: correlations

->Strong correlations between PageRank and CheiRank

->Strong correlation between moves which open many possibilities of new moves and moves that can follow many other moves.

->However, the symmetry is far from exact.

->Correlation less strong for largest network

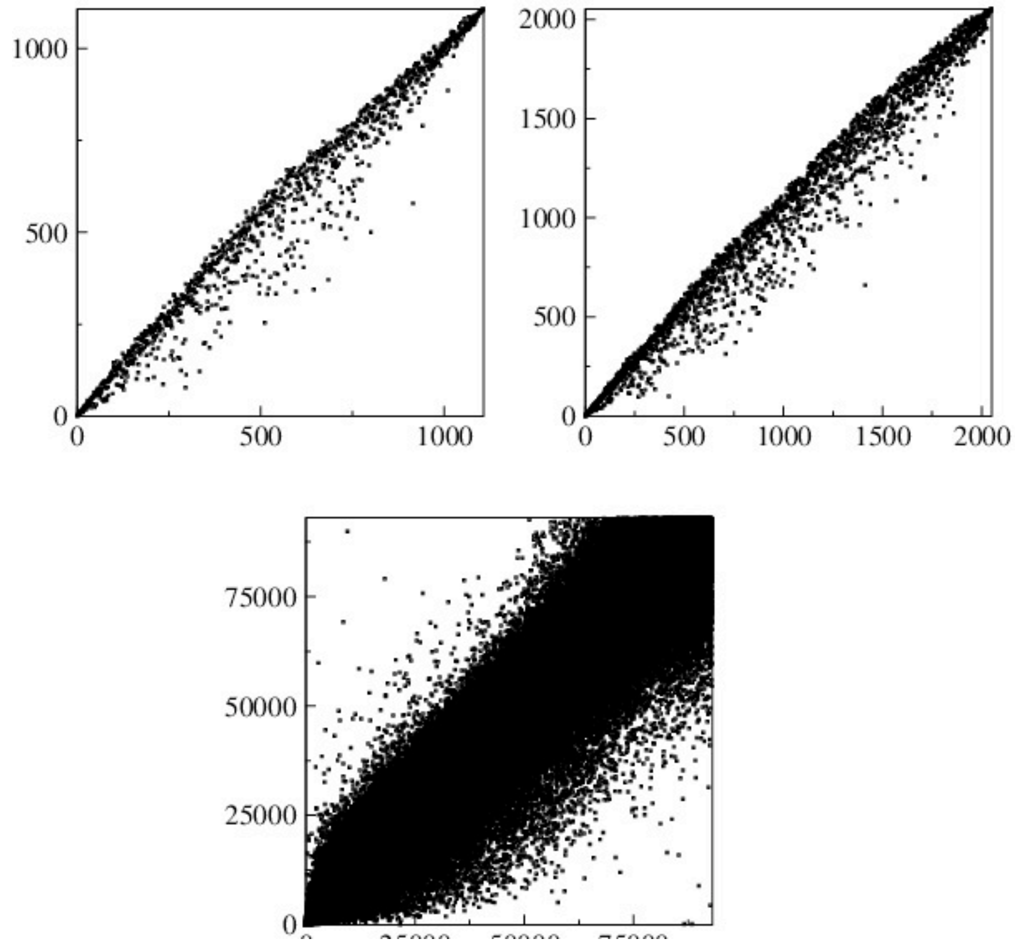
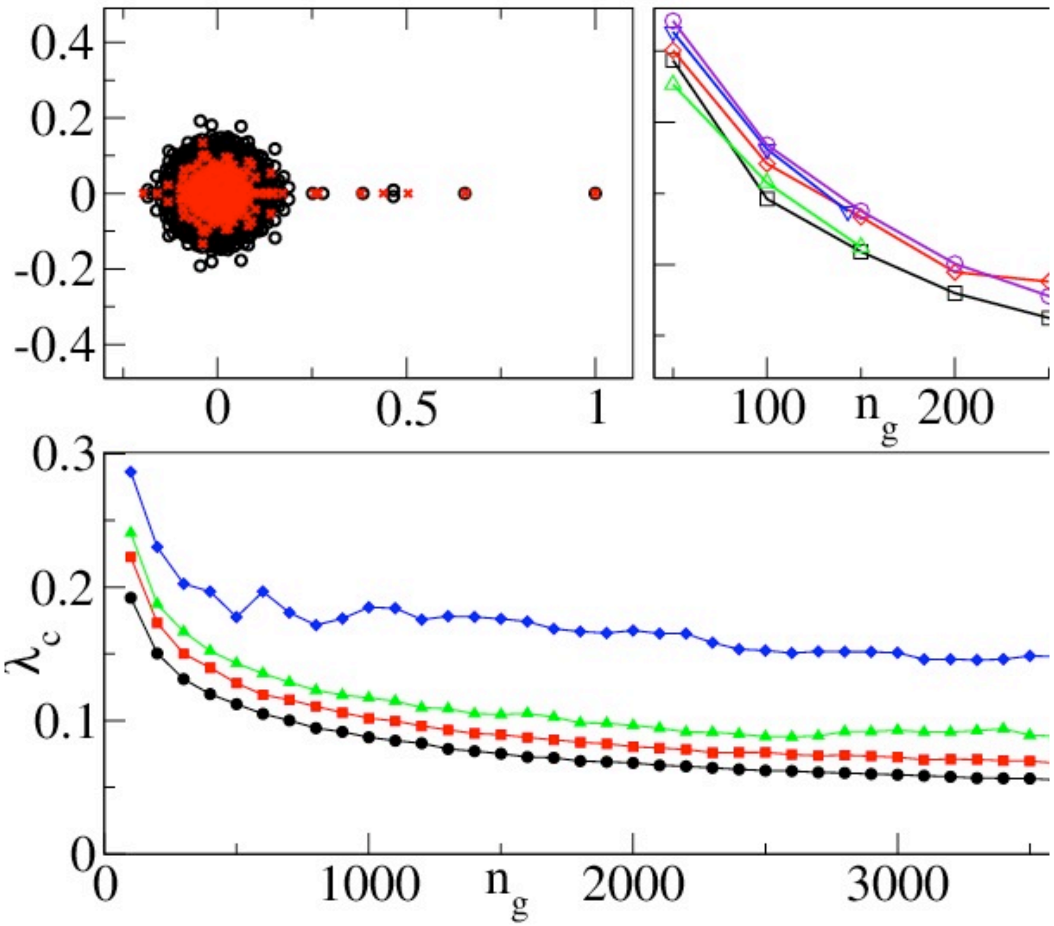


Figure: K* vs K where K (resp. K*) is the rank of a vertex when ordered according to PageRank vector (resp CheiRank) for the three networks (sizes 1107, 2051, 193995)

Spectrum of the Google matrix I

->For WWW the spectrum is spread inside the unit circle, no gap between first eigenvalue and the bulk
->Here huge gap like in well-connected networks, with few isolated communities (cf lexical networks).

.



Top left: eigenvalues of G in the complex plane for two databases
Bottom: value such that from top to bottom 99%, 95%, 90%, 80% of eigenvalues are smaller in modulus for amateur games.
Top right: value for 80% of eigenvalues for our 5 databases.

Spectrum of the Google matrix II

->For second and third networks, still gap between the first eigenvalue and next ones
->Radius of the bulk of eigenvalues changes with size of network
->More structure in the networks with larger plaquettes which disambiguate the different game paths and should make more visible the communities of moves

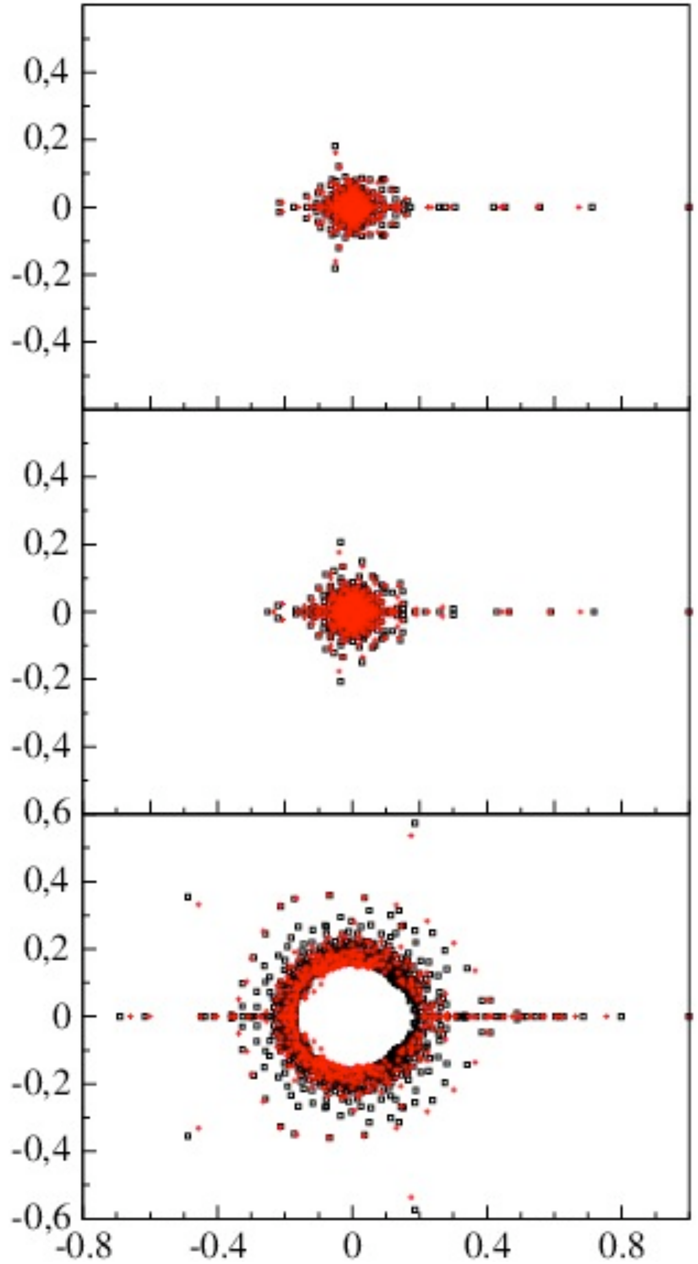


Figure: Eigenvalues of G in the complex plane for the networks with 1107, 2051 and 193995 nodes

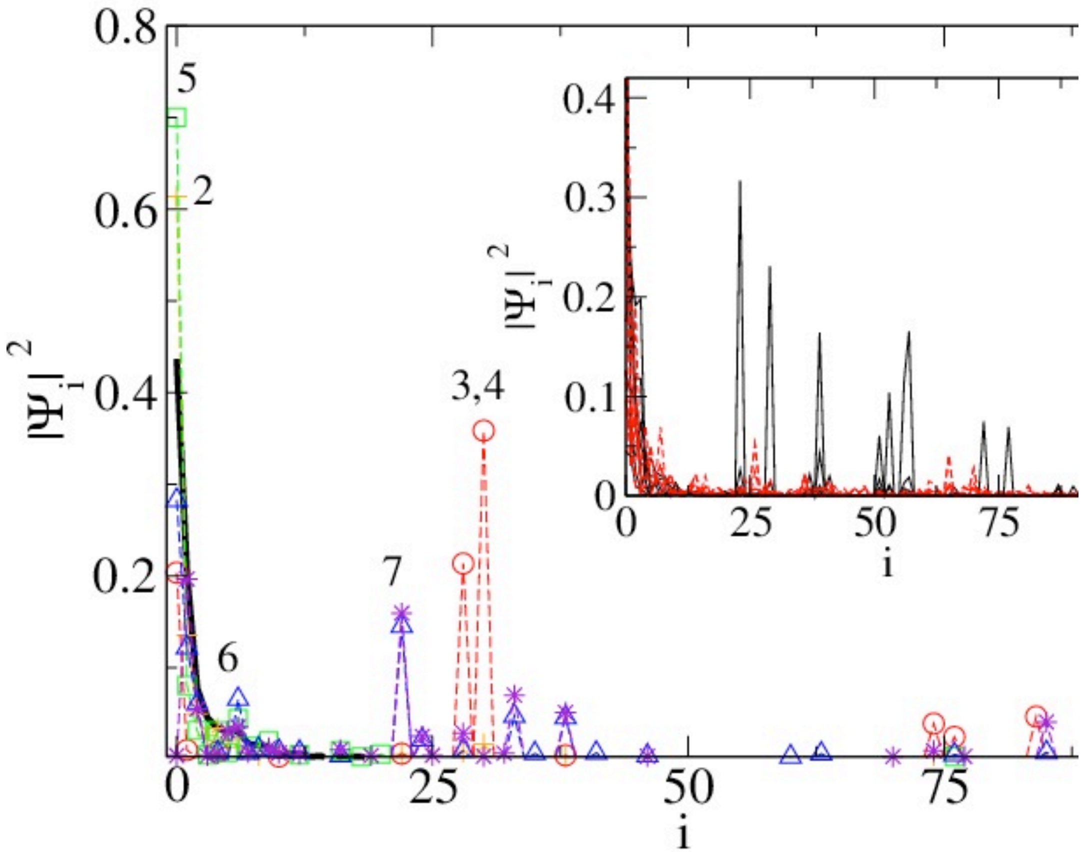
Eigenvectors of the Google matrix I

->Next to leading eigenvalues are important, may indicate the presence of communities of moves with common features.

->The distribution of the first 7 eigenvectors (Left) shows that they are concentrated on particular sets of moves different for each vector.

->eigenvectors are different for different tournaments and from professional to amateur

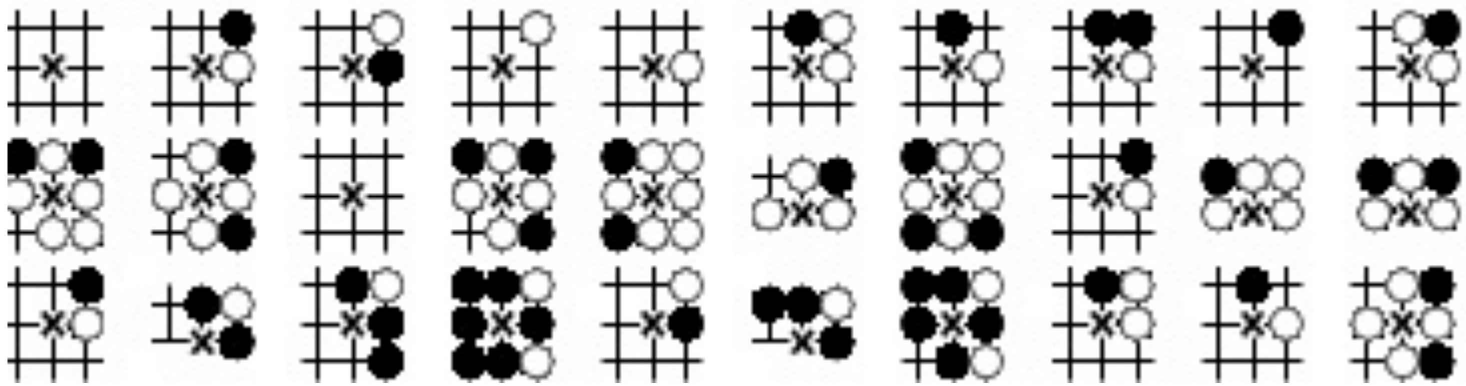
->much less peaked for randomized network



Moduli squared of the right eigenvectors of the 7 largest eigenvalues of G (network with 1107 vertices). Inset: real games (black) vs random network (red)

Connection with tactical sequences

->First eigenvector is mainly localized on most frequent moves
->Third one is localized on moves describing captures of the opponent's stones, and part of it singles out the well-known situation of *ko* (``eternity"), where players repeat captures alternately.
->The 7th eigenvector singles out moves which appear to protect an isolated stone by connecting it with a chain.



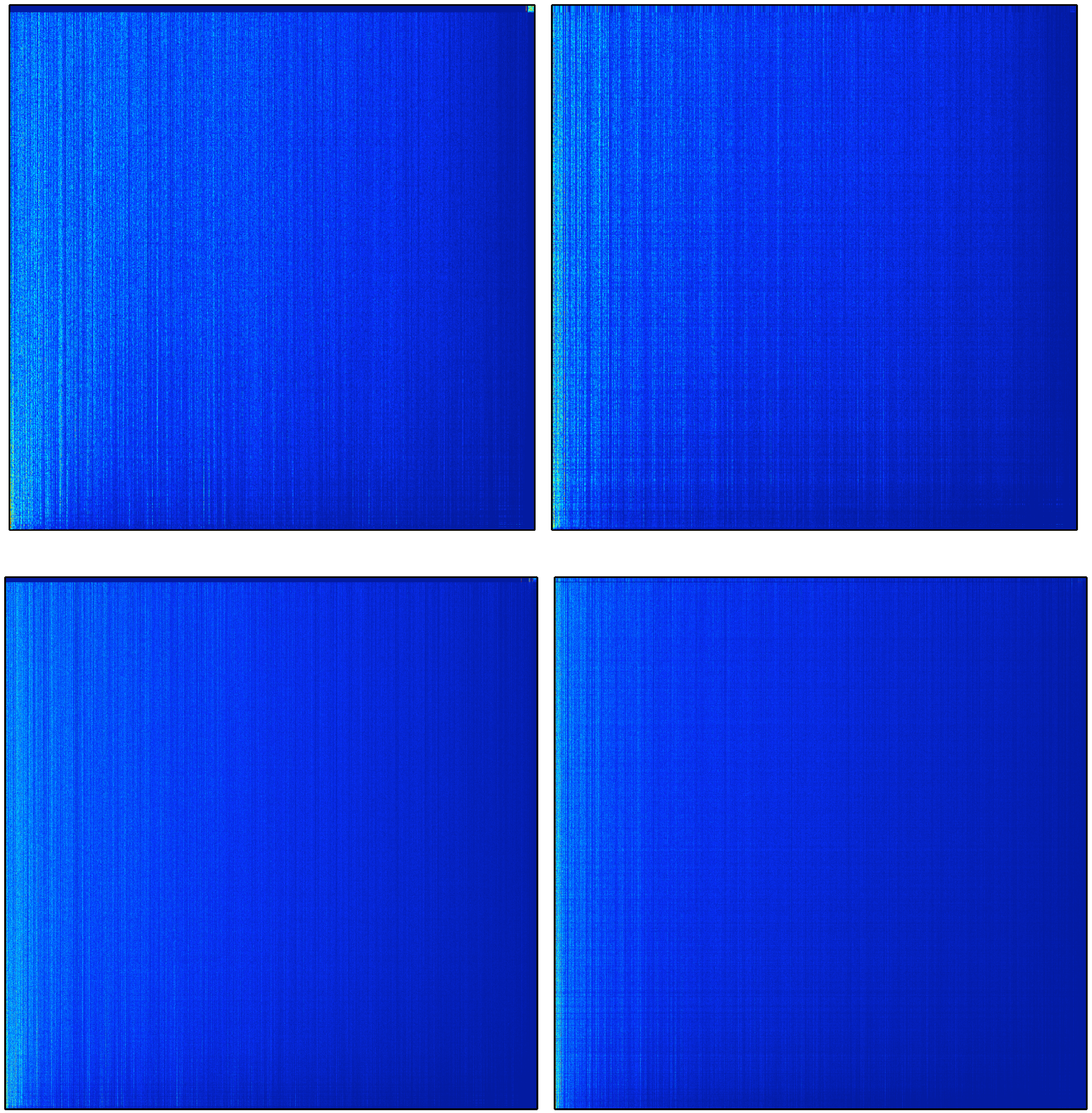
Moves corresponding to the 10 largest entries of right eigenvectors of G for first eigenvalues (PageRank)(top), third one (middle) and seventh one (bottom), Network with 1107 vertices.

Eigenvectors correlations

->Eigenvectors of network of size 1107 (top) and 2051(down)

->Left (right) One line: one eigenvector in the order of PageRank (CheiRank)

->Correlations visible, not necessarily related to PageRank

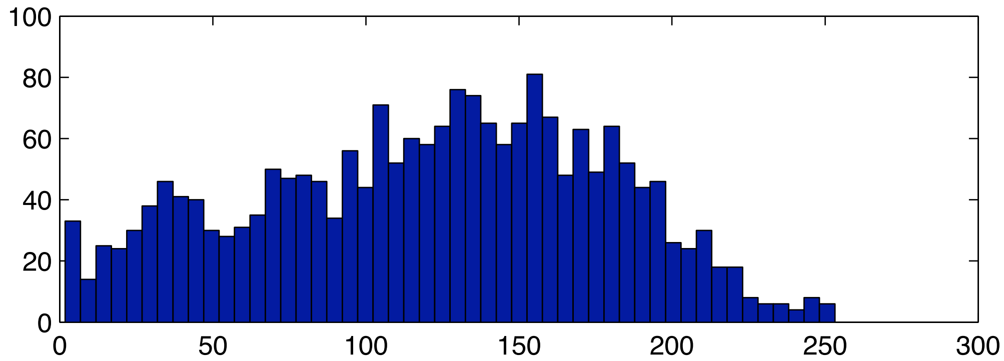


Eigenvectors localization

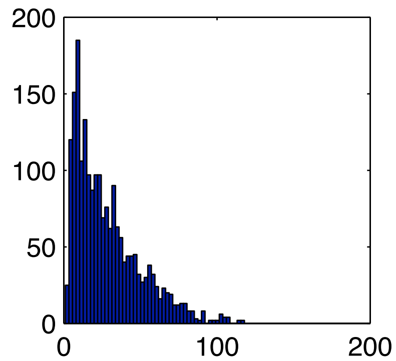
->Inverse participation ratio: measures the spreading of eigenvectors $(\Sigma_i |\backslash P_i|^2/\Sigma_i |\backslash P_i|^4)$

->Large dispersion for G (top)

-> Lower dispersion for G with links inverted (bottom)



Network of size 193995 (diamonds)



Eigenvectors of the Google matrix II

->More complicated groups of moves can be seen in eigenvectors of larger networks

->Systematic method of grouping them: by antecedent, by correlations between eigenvectors.

Figures: eigenvector for network of size 2051 (bottom) and 193995 (right)

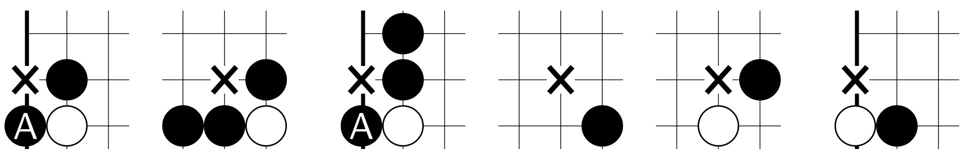


Figure 1: PageRank index : 27, 28, 52, 9, 6, 32,

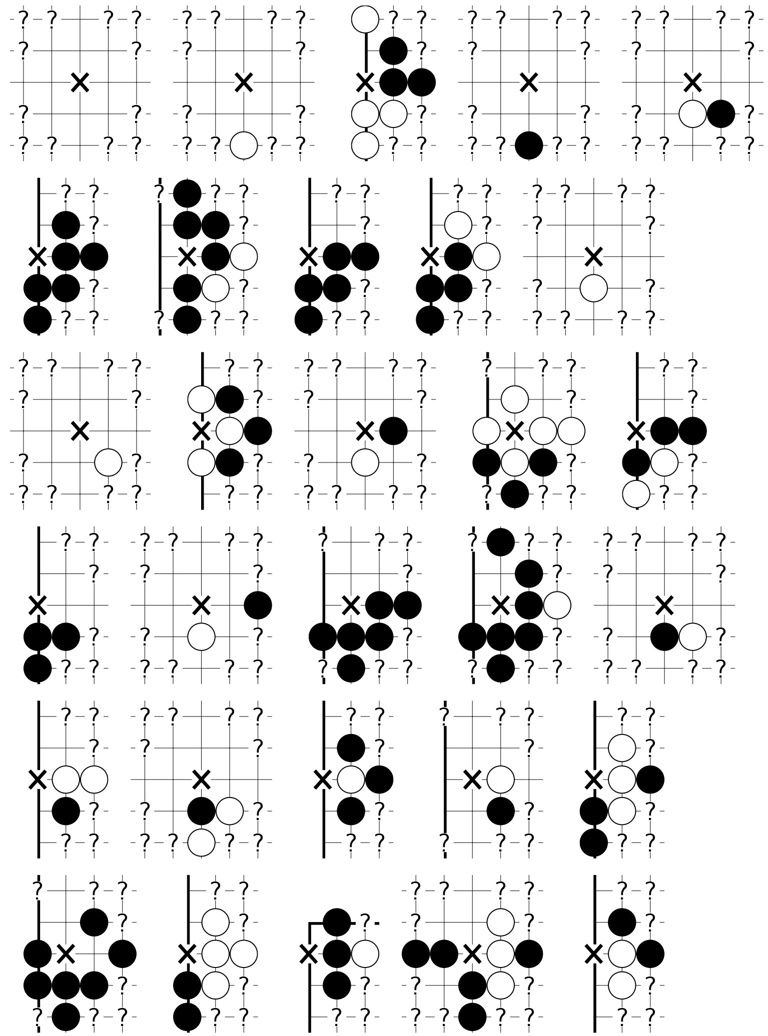
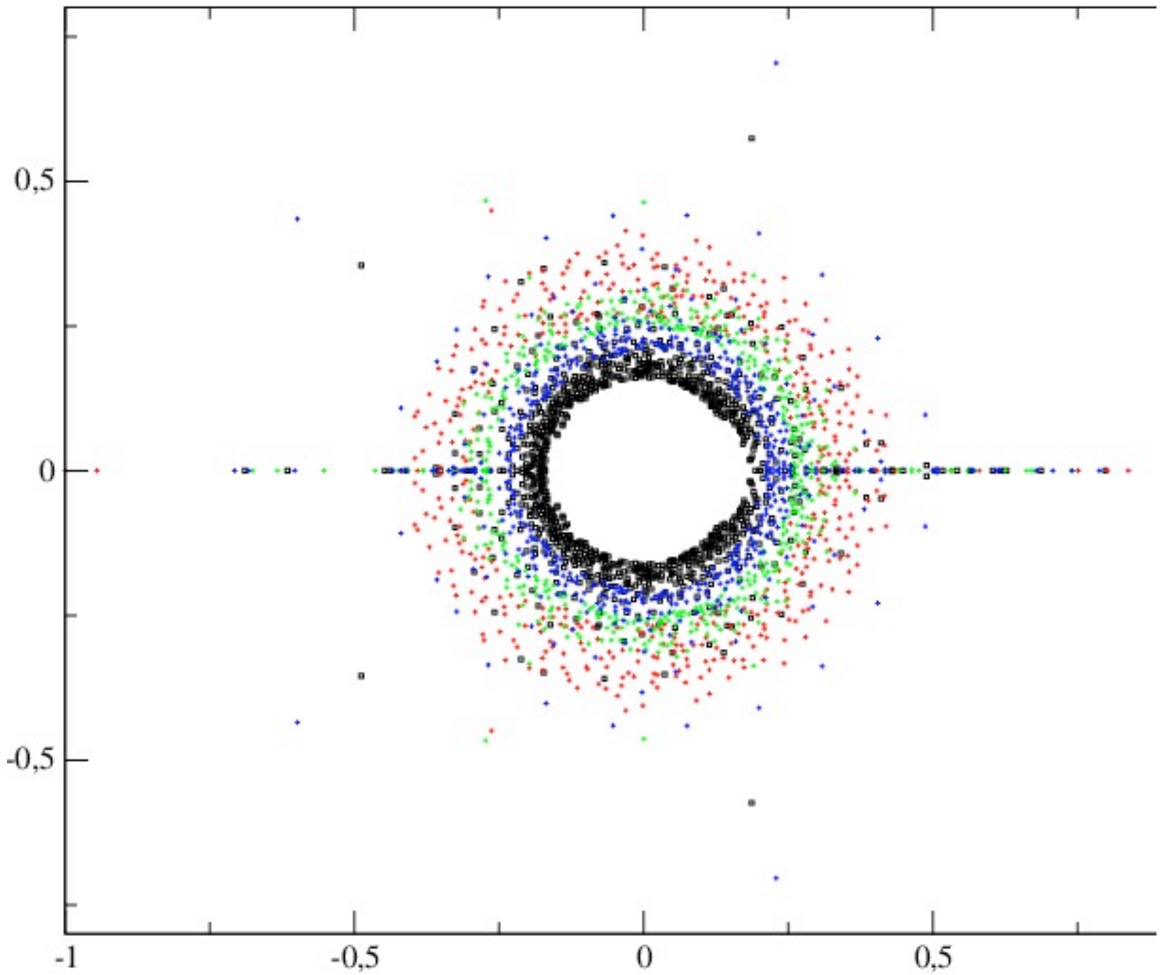


Figure 1: PageRank index : 0, 2, 10726, 1, 3, 35652, 63829, 56615, 45588, 6, 7, 144, 9, 126, 29, 63846, 10, 85819, 75486, 16, 14, 4, 21, 15, 1216, 77223, 1545, 35403, 24208, 22,

Networks for different game phases

->One can separate the games into beginning, middle, and end
->The three networks are different, with markedly different spectra and eigenvectors

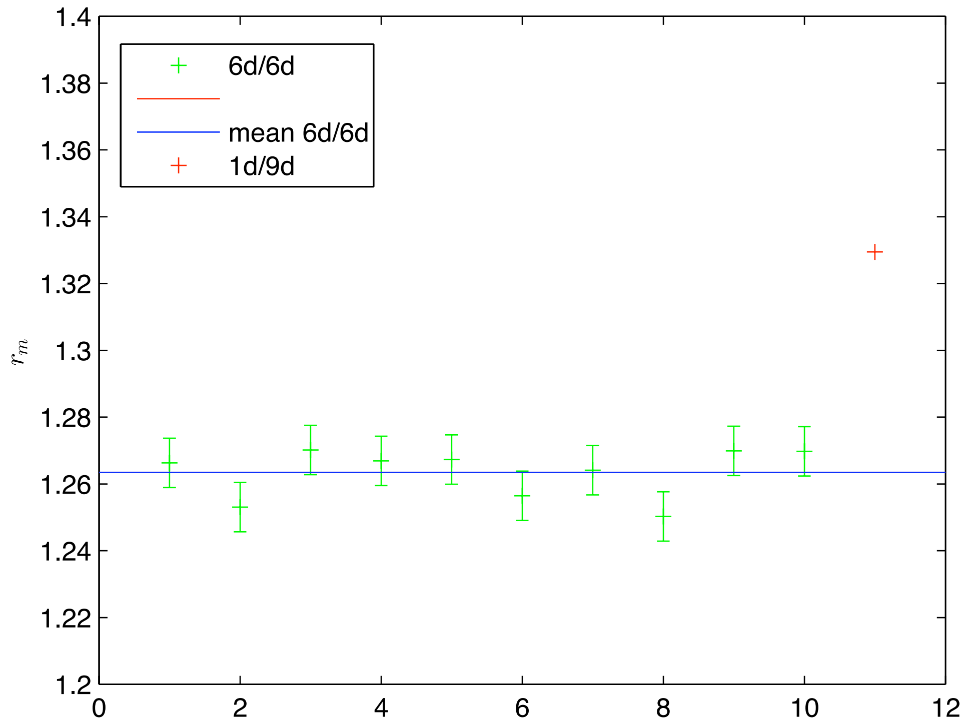
Figure: spectrum for all moves (black), 50 first moves (red), middle 50 (blue) and last 50 (green), Network with 193995 vertices.



Networks for different levels of play

->One can separate the players by their levels (dans)
->Differences can be seen between the moves of these players at the network level

Figure: statistical difference between nodes outdegrees for 1dan/9dan and several sets of 6dans/6dans Network with 193995 vertices.



Conclusion

->we have studied the game of go, one of the most ancient and complex board games, from a complex network perspective.

->We have defined a proper categorization of moves taking into account the local environment, and shown that in this Case Zipf's law emerges from data taken from real games

->Differences between professional and amateur games, different level of amateurs, or phases of the game.

->Certain eigenvectors are localized on specific groups of moves which correspond to different strategies.

->the point of view developed should allow to better modelize such games and could also help to design simulators which could in the future beat good human players.

->Our approach could be used for other types of games, and in parallel shed light on the human decision making process.