The game of go as a complex network

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

NADINE project REVIEW 14 November 2013, Toulouse

Networks

->Nadine project: new tools for directed network structure analysis

->Important examples from recent technological developments: World Wide Web, social networks...

->But network theory 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

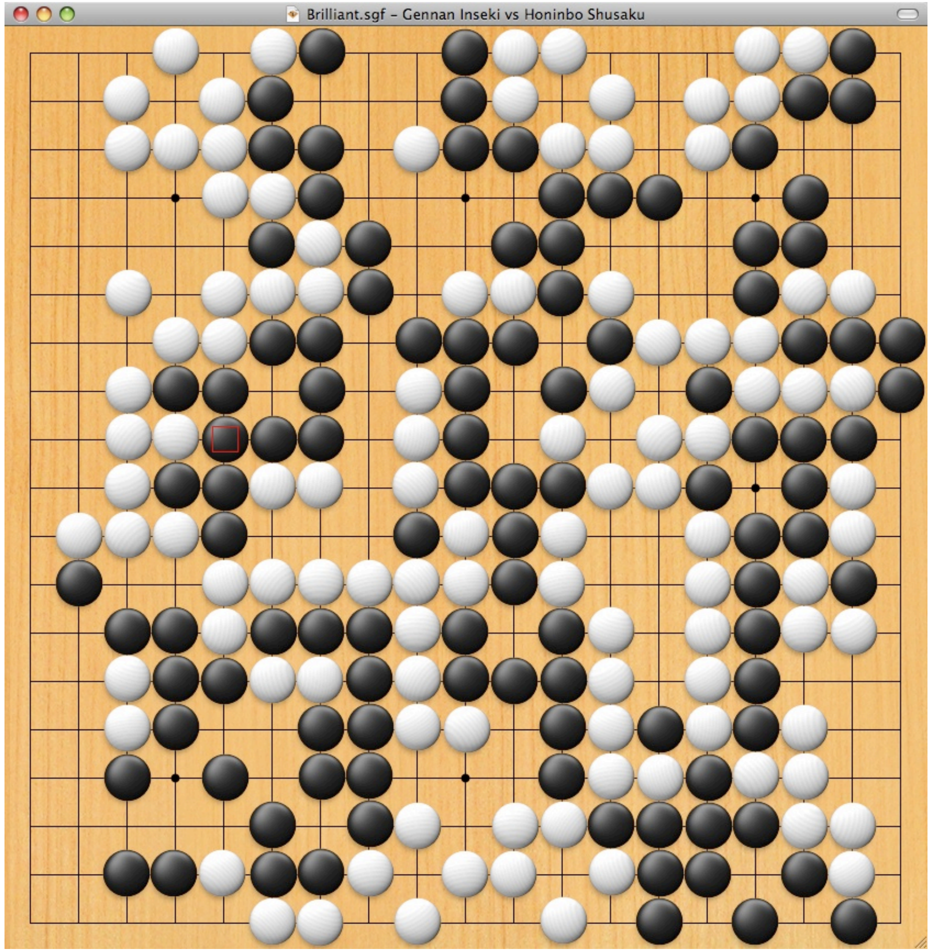
->White and black stones alternatively put at intersections of 19 x19 lines

->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:

->To increase statistics and compare with professional tournaments, 135 000 amateur games also used.

->Level of players is known

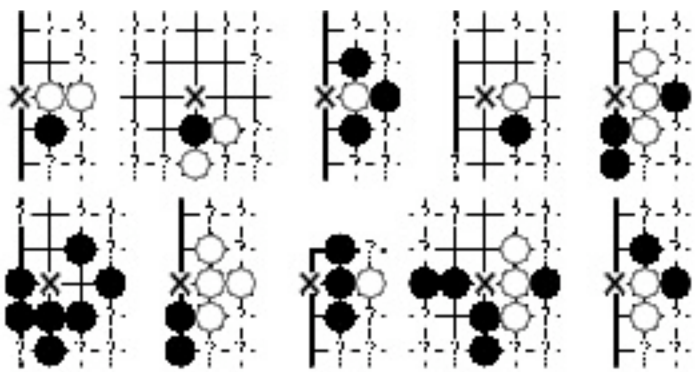
Vertices of the networks

->"plaquette" : three sizes: 1) 3x3 intersections, 2) 3x3 intersections with atari status of nearest neighbours, 3) diamond of 3 x3 +4 intersections

->We identify plaquettes related by symmetry or with color swapped

->Respectively 1107, 2051 and 193995 nonequivalent plaquettes with empty centers

->vertices of our network



Links of the network

->we connect vertices corresponding to moves a and b if b follows a in a game at a distance < d=5

->Sequences of moves follow Zipf's law (cf languages) Exponent decreases as longer sequences reflect individual strategies

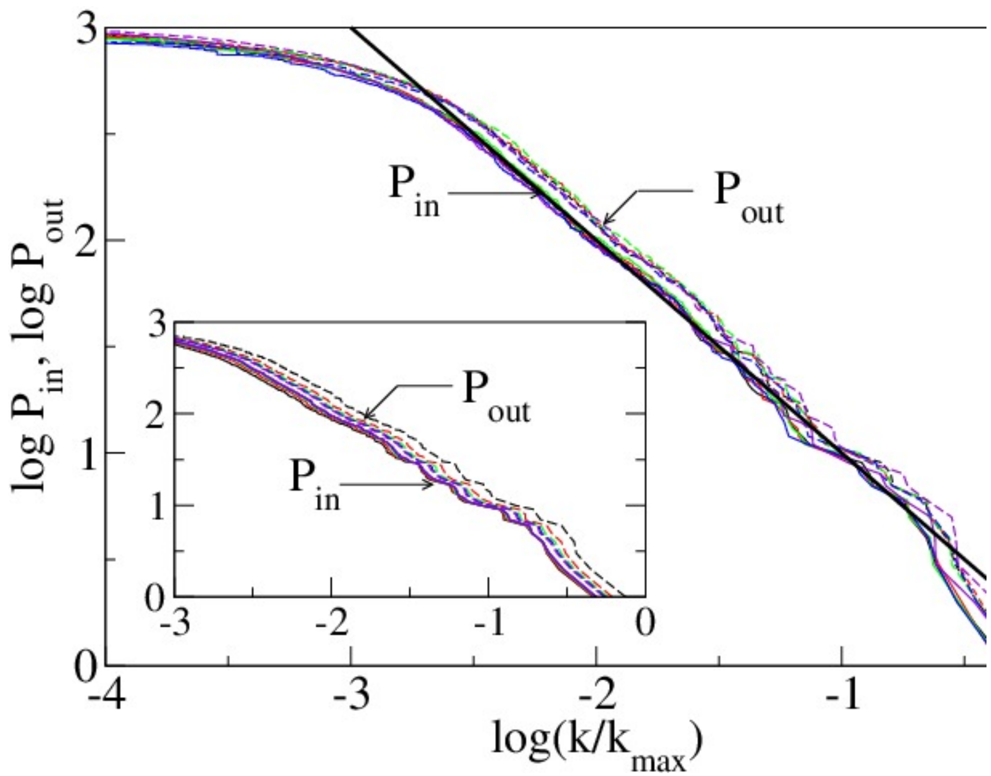
->amateur database departs from all professional ones, playing more often at shorter distances

Links distribution

->Tails of link distributions very close to power-law with exponent 1.0 for the integrated distribution.

->network displays the scale-free property

->symmetry between ingoing and outgoing links is a peculiarity of this network



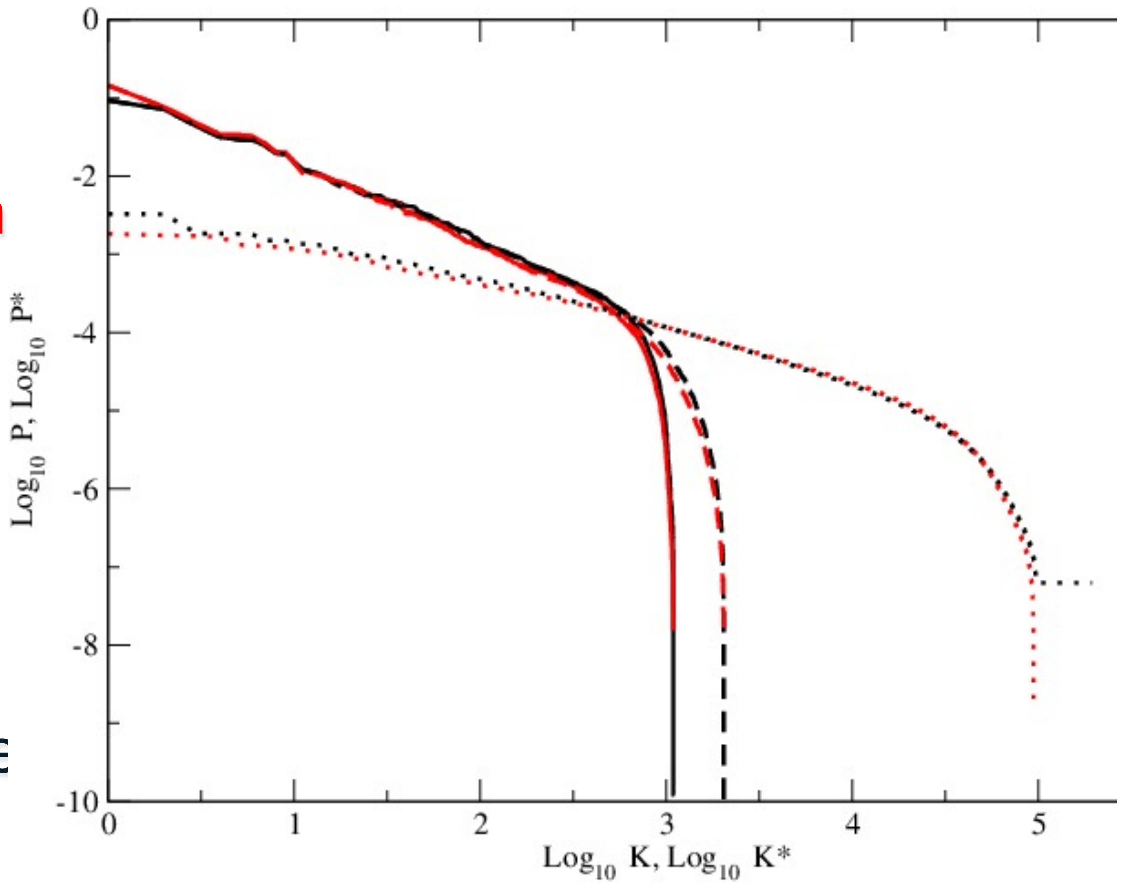
Normalized integrated distribution of links for d=5
Thick solid line is y=-x.
Inset:different values of d

Ranking vectors

->PageRank: ingoing links
->CheiRank: outgoing links

->Ranking vectors follow an algebraic law

->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

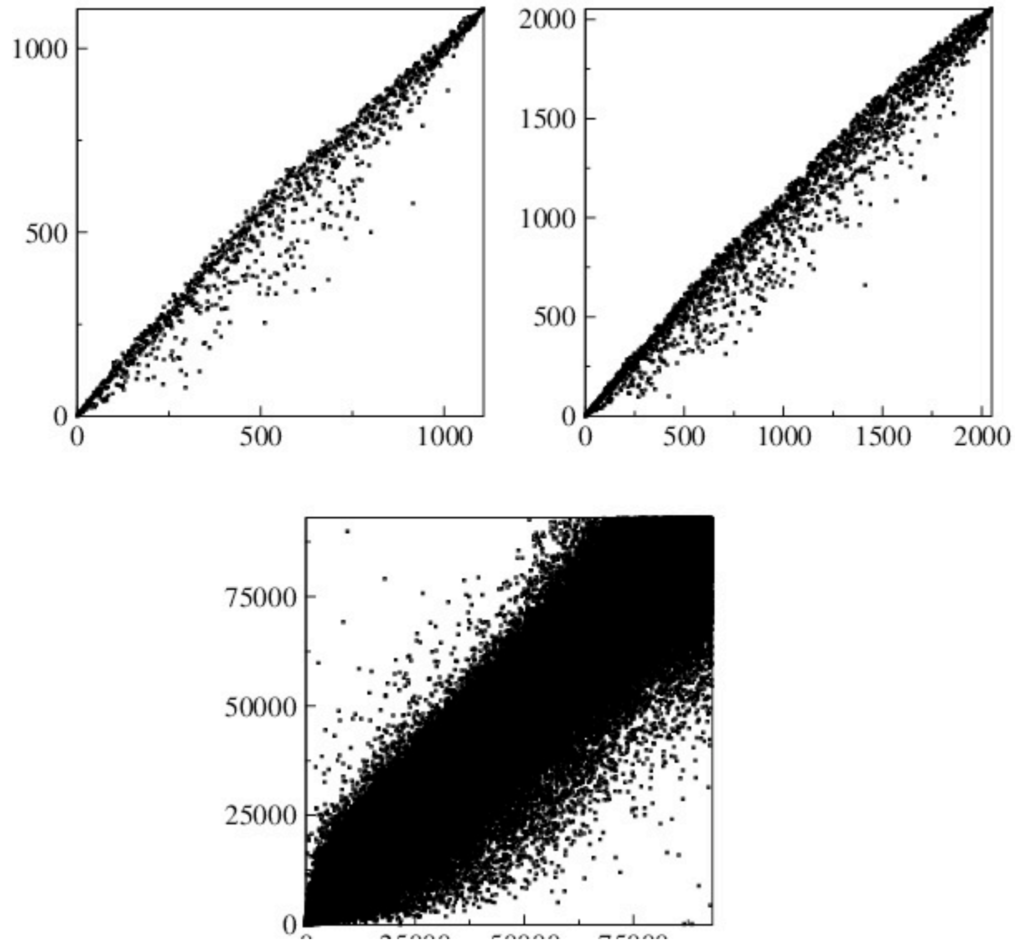


Figure: K* vs K where K (resp. K*) is the rank of a vertex when ordered according to PageRank vector (resp CheiRank) for the three networks (sizes 1107, 2051, 193995)

Spectrum of the Google matrix

->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).

->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

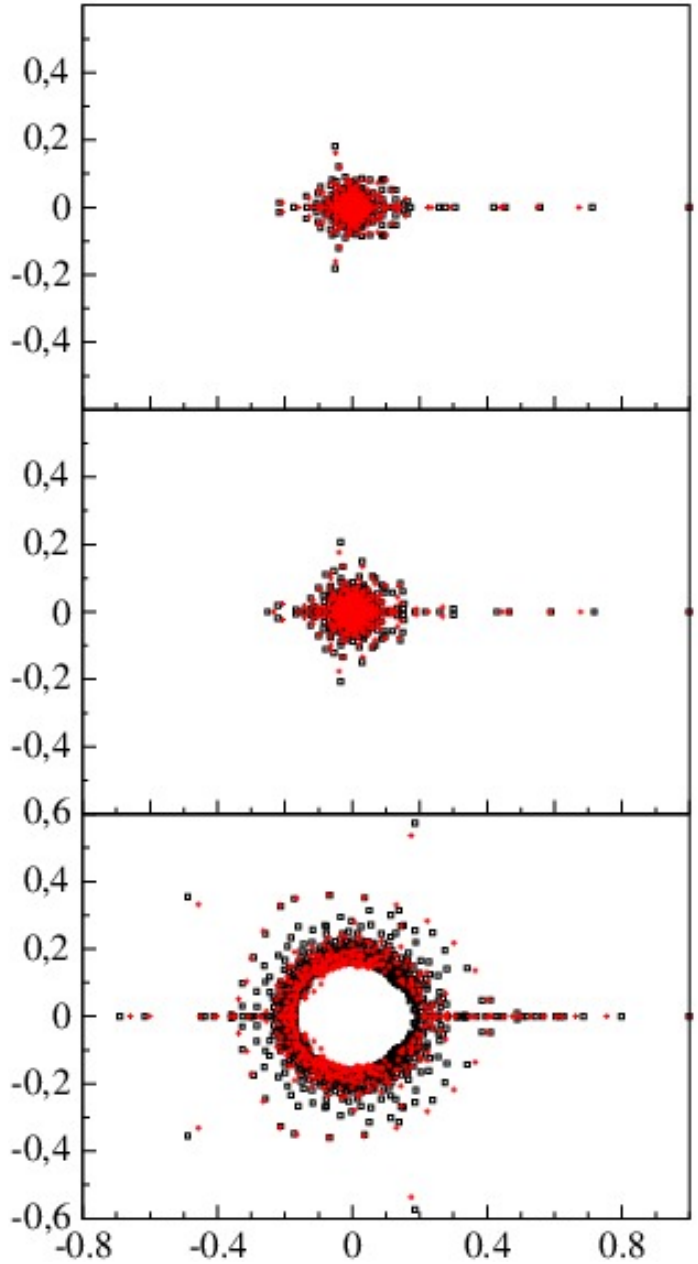
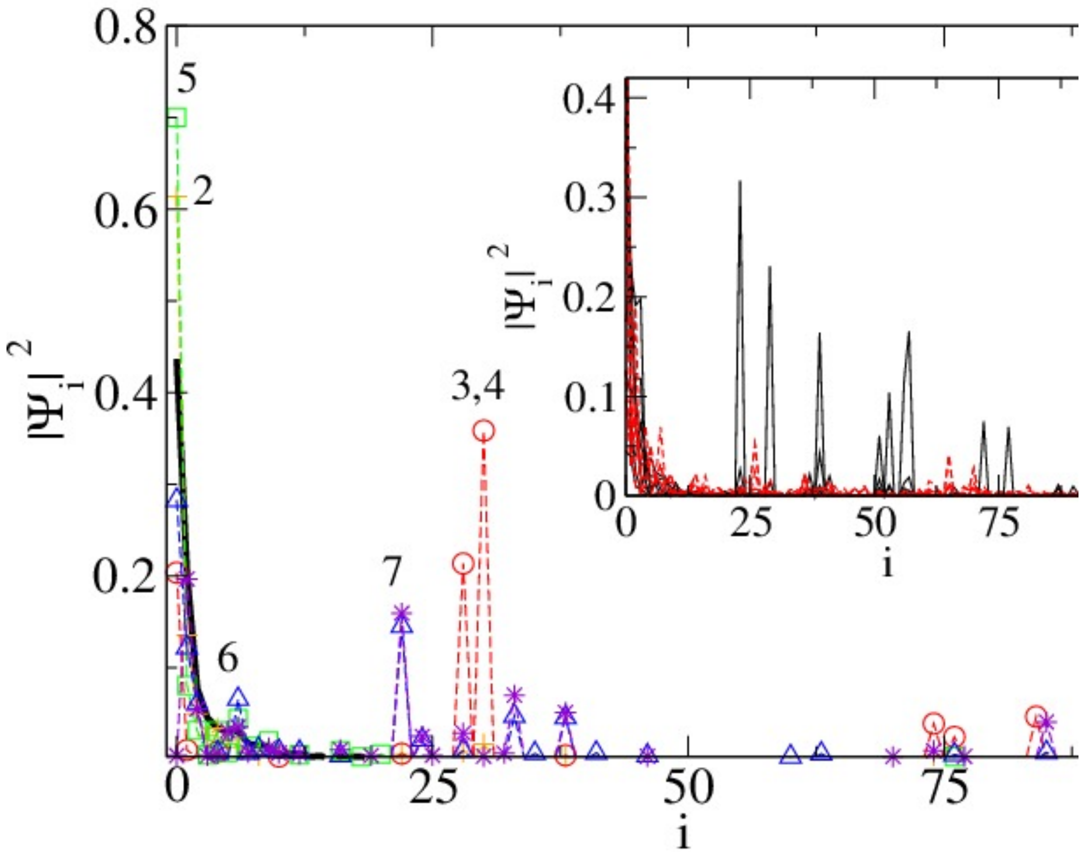


Figure: Eigenvalues of G in the complex plane for the networks with 1107, 2051 and 193995 nodes

Eigenvectors of the Google matrix I

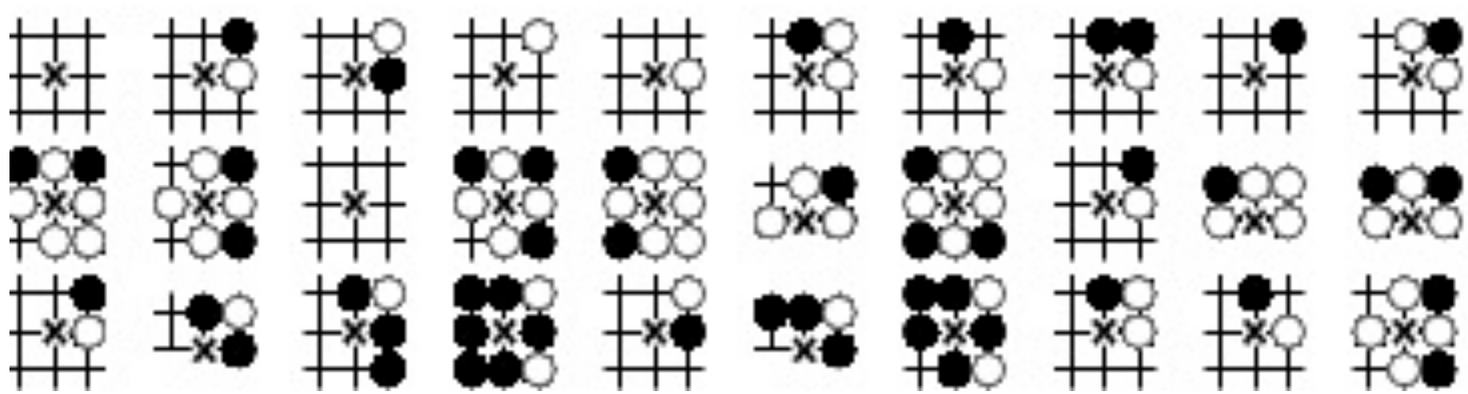
->Next to leading eigenvalues are important, may indicate the presence of communities of moves with common features.
->The distribution of the first 7 eigenvectors (Left) shows that they are concentrated on particular sets of moves different for each vector.
->eigenvectors are different for different 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 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)

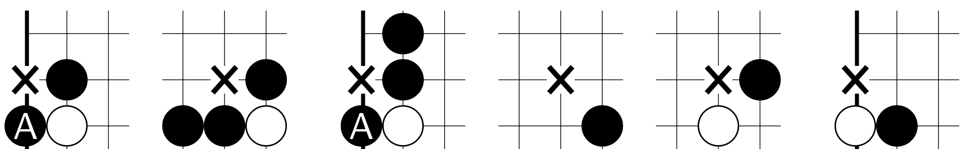


Figure 1: PageRank index : 27, 28, 52, 9, 6, 32,

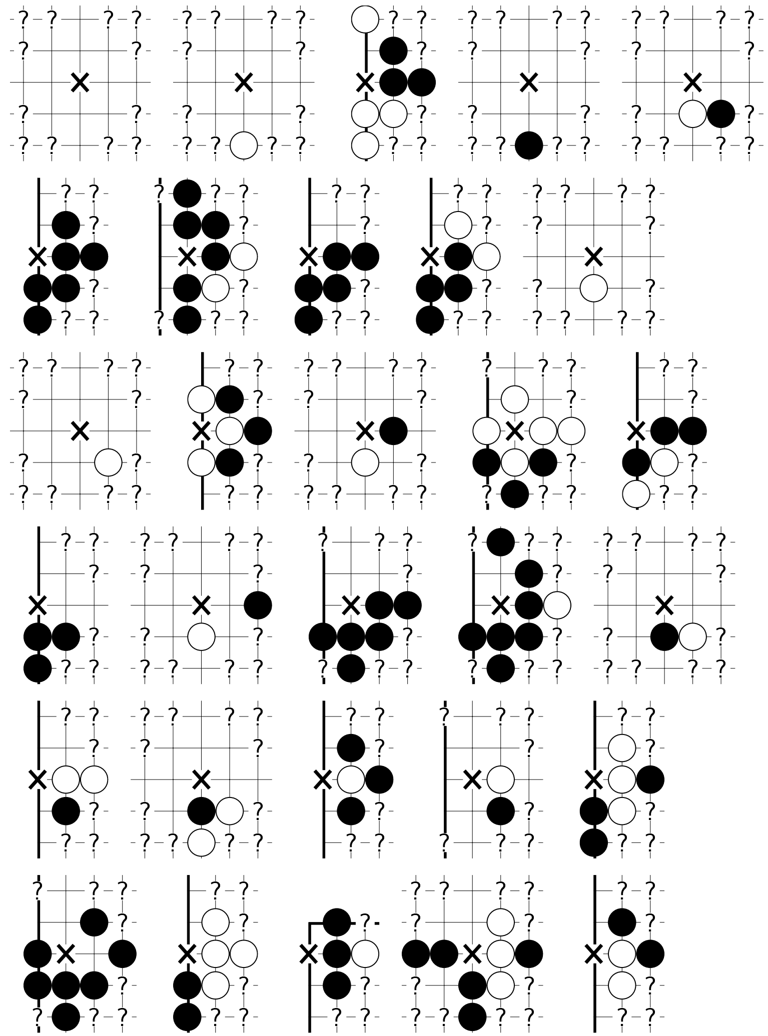
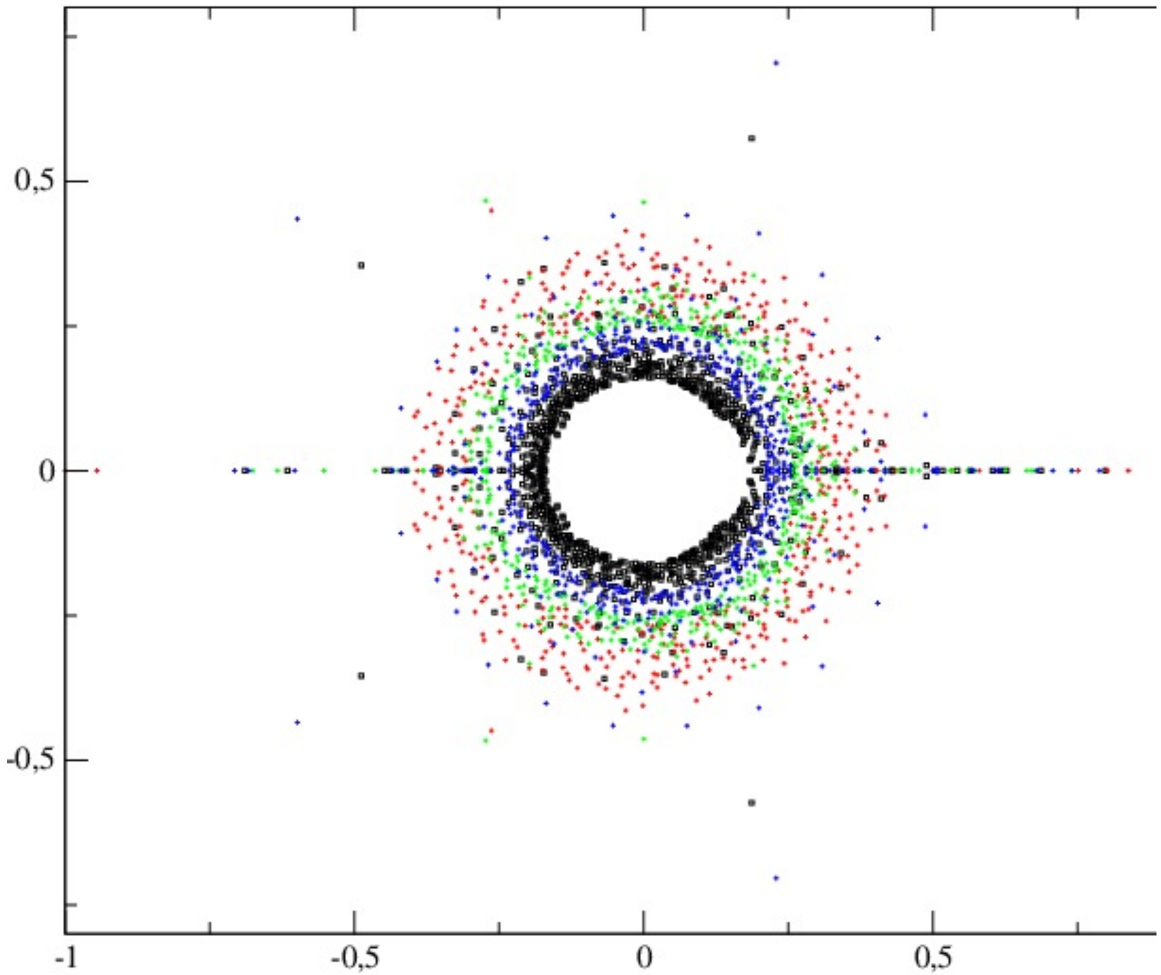


Figure 1: PageRank index : 0, 2, 10726, 1, 3, 35652, 63829, 56615, 45588, 6, 7, 144, 9, 126, 29, 63846, 10, 85819, 75486, 16, 14, 4, 21, 15, 1216, 77223, 1545, 35403, 24208, 22,

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

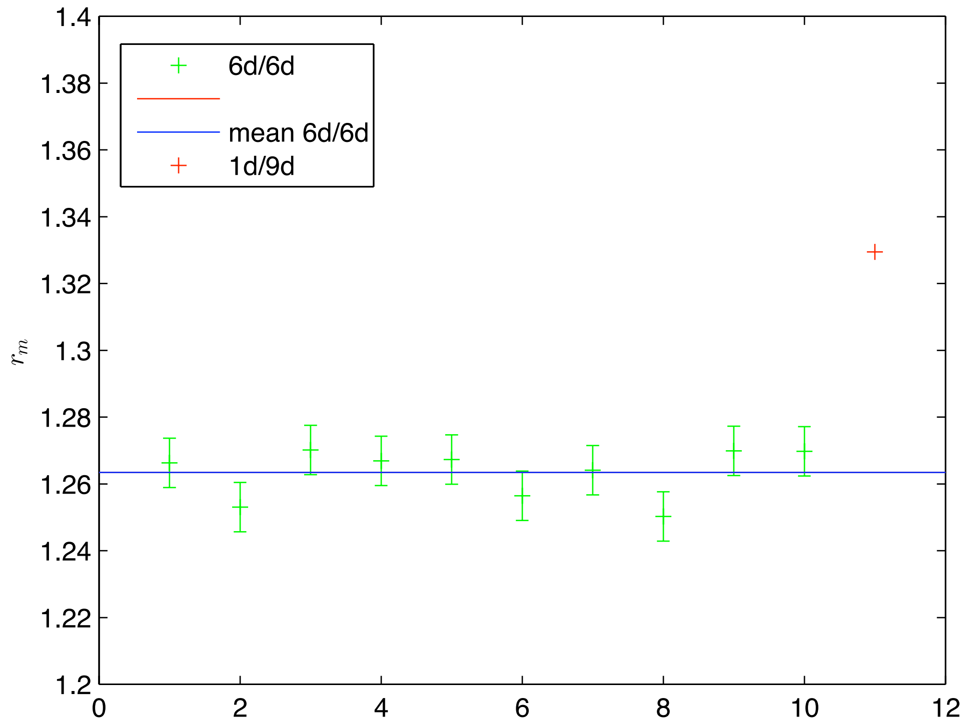


Figure: statistical difference between nodes outdegrees for 1dan/9dan and several sets of 6dans/6dans Network with 193995 vertices.

Conclusion

-> We have built a complex network which describes the game of go, one of the most ancient and complex board games.

-> Network structure analyzed with Nadine tools show 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.