The game of go as a complex network

Bertrand Georgeot, Olivier Giraud, Vivek Kandiah

supported by EC FET Open project NADINE

B.G. and O. Giraud, Europhysics Letters **97** 68002 (2012)

Quantware group
Laboratoire de Physique Théorique, IRSAMC, UMR 5152 du CNRS
Université Paul Sabatier, Toulouse

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...

Games

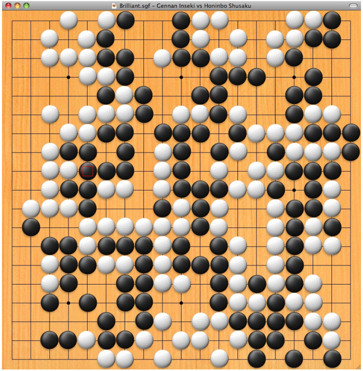
- Network theory never applied to games
- Games represent a privileged approach to human decision-making
- Can be very difficult to modelize or simulate

⟹ 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.

Goban

Rules of go

- White and black stones alternatively put at intersections of 19$\times$ 19 lines
- Stones without liberties are removed
- Handicap stones can be placed
- Aim of the game: construct protected territories
- total number of legal positions $\sim 10^{171}$, compared to $\sim 10^{50}$ for chess



Databases

- We use databases of expert 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)
- To increase statistics and compare with professional tournaments, 4000 amateur games were also used.

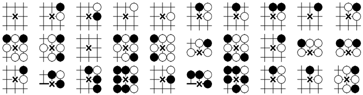
Vertices of the network

"plaquette" $\Rightarrow$ square of 3×3 intersections

- We identify plaquettes related by symmetry
- We identify plaquettes with colors swapped

$\Longrightarrow$ 1107 nonequivalent plaquettes with empty centers

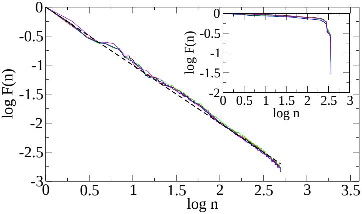
$\Longrightarrow$ vertices of our network



Examples of plaquettes

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 an exponent ≈ 1.06



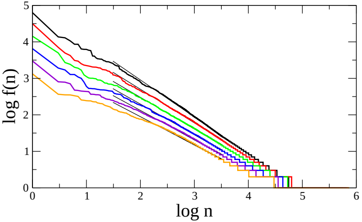
Normalized integrated frequency distribution of 1107 moves. Thick dashed line is $y = -x$. Inset: same for positions on the board

Sequences of moves

- we connect vertices corresponding to moves *a* and *b* if *b* follows *a* in a game at a distance $\leq d$.
- Each choice of *d* defines a different network.
- Left: frequency distribution for sequences of the 1107 moves with $d = 4$. Algebraic decrease visible, exponent from ≈ 1 (short sequences) to ≈ 0.7 (long sequences).

$\Longrightarrow$ Sequences of moves follow Zipf's law (cf languages)

$\Longrightarrow$ Exponent decreases as longer sequences reflect individual strategies



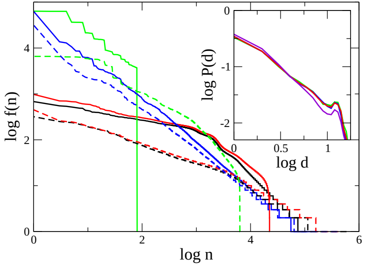
Integrated frequency distribution of sequences of moves $f(n)$ for (from top to bottom) two to seven successive moves (all databases together), plotted against the ranks of the moves.

Sequences of moves

Four possible definitions:

- C1: positions on the board, *b* follows *a* if *b* is played immediately after *a*
- C2:positions on the board, *b* follows *a* if *b* is played after *a* at distance $d = 4$
- C3: sequence of vectors between successive positions with $d = 4$
- C4: as before

$\Longrightarrow$ move sequences, even long ones, are well hierarchized by our initial definition
$\Longrightarrow$ amateur database departs from all professional ones, playing more often at shorter distances



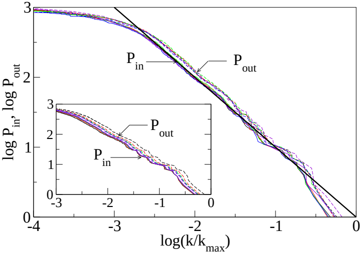
Integrated frequency distribution of sequences of moves for two (continuous) and three (dashed lines) successive moves, cases C1 (black), C2 (red), C3 (green), C4 (blue). Inset: distribution of distances between moves $P(d)$. All professional tournaments are different from amateur games.

Link distributions

- Tails of link distributions very close to a power-law $1/k^{\gamma}$ with exponent $\gamma = 1.0$ for the integrated distribution.
- The results are stable in the sense that the exponent does not depend on the database considered.

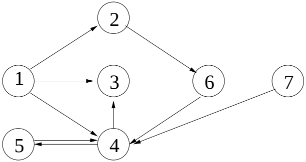
$\Longrightarrow$ network displays the scale-free property

$\Longrightarrow$ symmetry between ingoing and outgoing links is a peculiarity of this network



Normalized integrated distribution of ingoing links $P_{\rm in}$ (solid) and outgoing links $P_{\rm out}$ (dashed), Thick solid line is $y = -x$.
Inset: $P_{\rm in}$ (solid curves) and $P_{\rm out}$ (dashed curves), $d = 2$ (black), 3 (red), 4 (green), 5 (blue) and 6 (violet).

Directed network: Google algorithm



Weighted adjacency matrix

$$H = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Ranking pages $\{1, \ldots, N\}$ according to their importance.
PageRank vector $\mathbf{p}$ = stationary vector of H:

Computation of PageRank

$\mathbf{p} = H\mathbf{p} \Rightarrow \mathbf{p}$= stationary vector of H:
can be computed by iteration of H.

To remove convergence problems:

Replace columns of 0 (dangling nodes) by $\frac{1}{N}$: $H \rightarrow$ matrix S

In our example, $H = \begin{pmatrix} 0 & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{7} & 0 & 1 & 1 & 1 \\ 0 & 0 & \frac{1}{7} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 1 & \frac{1}{7} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{7} & 0 & 0 & 0 & 0 \end{pmatrix}$.

To remove degeneracies of the eigenvalue 1, replace S by

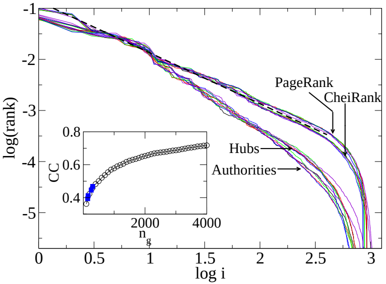
$$G = \alpha S + (1 - \alpha)\frac{1}{N}$$

Ranking vectors

- The PageRank algorithm gives the PageRank vector, with amplitudes p_i, with $0 \leq p_i \leq 1$
- PageRank is based on ingoing links
- One can define a similar vector based on outgoing links (CheiRank)
- HITS algorithm: Authorities (ingoing links) and Hubs (outgoing links)
- Other eigenvalues and eigenvectors of G reflect the structure of the network

Ranking vectors

- Clustering coefficient detects local connected clusters.
- Here depends on the number of games n_g included, but almost not on the database.
- For large n_g, it goes to an asymptotic value which seems larger than 0.7 (higher CC than WWW ≈ 0.11)
- Ranking vectors follow an algebraic law
- Symmetry between distributions of ranking vectors based on ingoing links and outgoing links.



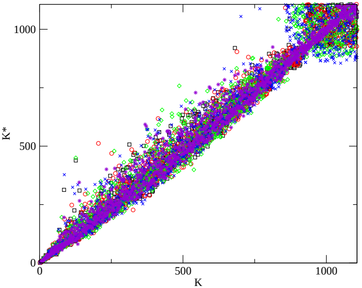
Ranking vectors of G. Top bundle: PageRank. Second bundle: CheiRank. Third bundle: Hubs. Fourth bundle: Authorities. Straight dashed line is $y = -x$. Inset: Clustering coefficient as a function of the number of games n_g included to construct the network; blue squares: professional tournaments; circles: amateur games.

PageRank vs CheiRank

- Left: correlation between the PageRank and the CheiRank for the five databases considered.
- Strong correlation between these rankings based respectively upon ingoing and outgoing links.

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

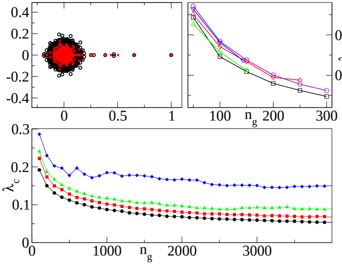
⟹ However, the symmetry is far from exact



K* vs K where K (resp. K*) is the rank of a vertex when ordered according to PageRank vector (resp CheiRank) for amateur (violet stars) and professional (other) databases.

Spectrum of the Google matrix

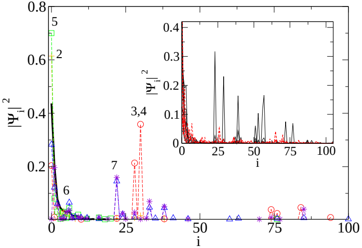
- For WWW the spectrum is spread inside the unit circle, no gap between first eigenvalue and the bulk
- Here huge gap between the first eigenvalue and next ones $\Longrightarrow$ well-connected network, few isolated communities (cf lexical networks).
- Radius of the bulk of eigenvalues changes with number of games n_g $\Rightarrow$ As more games are taken into account, rare links appear which break the weakly coupled communities.



Top left: eigenvalues of G in the complex plane; black circles: Honinbo; red crosses: amateur. Bottom: λ_c such that from top to bottom 99%, 95%, 90%, 80% of eigenvalues λ verify $|\lambda| < \lambda_c$ for amateur games. Top right: λ_c for 80% of eigenvalues for our 5 databases.

Eigenvectors of the Google matrix

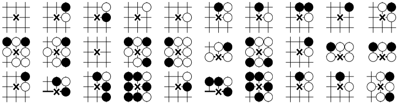
- Next to leading eigenvalues are important, as they indicate the presence of communities of moves which have common features.
- The distribution of the first 7 eigenvectors (Left) shows that they are concentrated on particular sets of moves different for each vector.
- eigenvectors are different for different tournaments and from professional to amateur
- much less peaked for randomized network



Moduli squared of the right eigenvectors associated with the 7 largest eigenvalues $|\lambda_1| = 1 > |\lambda_2| ... > |\lambda_7|$ of G (Honinbo database) for the first 100 moves in decreasing frequency.
Inset: Same for amateur database (black) and random network (red).

Connection with tactical sequences

- First eigenvector is mainly localized on the most frequent moves
- Third one is localized on moves describing captures of the opponent's stones, and part of them single 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 the10 largest entries of right eigenvectors of G for eigenvalues λ_1 (PageRank)(top), λ_3 (middle) and λ_7 (bottom), Honinbo database. Black is playing at the cross. Top line coincides with the 10 most frequent moves.

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 different tournaments.
- some peculiarities, such as a statistical symmetry between ingoing and outgoing links distributions
- Differences between professional tournaments and amateur games can be seen.
- Certain eigenvectors are localized on specific groups of moves which correspond to different strategies.

$\Longrightarrow$ the point of view developed in this paper should allow to better modelize such games

$\Longrightarrow$ could also help to design simulators which could in the future beat good human players.

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

$\Longrightarrow$ Future: larger plaquettes, comparison human/computers