# EXAMINING ONLINE POKER AS A SOCIAL NETWORK

YASHIT JAIN, VIDUSHI SOOD, PULKIT PAHUJA, [4]PROF. GURPREET KAUR
[1,2,3]Students, [4]Assistant Professor
Department of Computer science Engineering
MVSIT, Sonipat, India

ABSTRACT
Data in the form of approximately 5 million No Limit Texas Hold'em hand histories collected from the online poker website PokerStars.com is used to model a game playing social network. Each player represents a node in the network, and weighted, directed edges represent the flow of money from the source to the target node, where the weight is equal to the amount of money transferred, normalized by stakes. Properties of the network structure such as degree distribution and weighted clustering coefficient are examined across different levels of stakes, ranging from $100 buy-in games (blinds of $.50/$1) to $1,000 buy-in games (blinds of $5/$10). We also explore the evolution of the network over time, creating probabilistic models for node arrivals and edge initiations. These models are governed in part by the unique set of incentives for edge creation, where each node seeks inward edges, but tries to avoid outward edges. We noticed several interesting patterns common to all networks across stakes, and attempt to utilize that information in our probabilistic model of edge generation. Despite our best efforts, the standard model of preferential attachment observed in class still performs best.

INTRODUCTION

Of all the games played in a casino, poker is the most social. Rather than playing against the house, the game is played against a table full of other players. In 2003, the world of online poker exploded and millions of people flocked to websites such as PartyPoker.com and PokerStars.com to play the card game on the Internet. When thinking of poker as an online network, there are several factors that distinguish it from more traditional networks such as Facebook or MySpace. For example, people playing online poker can choose the level of stakes they play for. Those who simply play for fun may choose to play for lower stakes, while those playing more seriously, or perhaps even professionally, might wish to play for higher stakes. Different levels attract different player pools, and thus we might expect to observe differences in their network structures. Also, edges are signed and weighted, and the creation of edges is adversarial in nature. Players are incentivized to generate inward pointing edges, which represents winning money, and disincentivized to generate outward pointing edges, which represents losing money.

2. ANALYSIS

Our analysis of the world of online poker was divided into three parts: collecting data, developing an understanding of the networks and their structures, and building a model for their evolution.

2.1 Data Collection

All hands of poker played on reputable websites are publicly viewable, and several websites offer datamining services to players. For a fee, a player can download hundreds of thousands of hand histories from websites such as HandHQ.com. These hand histories can be stored in a database, and serve as information on an opponent's betting tendencies and inclinations. We contacted HandHQ customer support and requested a large sample of hand histories for research purposes. HandHQ replied and sent us approximately 5 million hand histories from PokerStars with obfuscated unique identifiers in place of usernames[2]. The histories contain hands played at
$.5/$1, $1/$2, $3/$6 and $5/$10 stakes. We used Python to process the hand histories and extract relevant information such as players, stakes, time and hand outcomes. Each of our hand results are normalized by the level of stakes played. In our data, we use the big blind as our currency unit, which is the number after the forward slash, e.g., one big blind is $1 at $.5/$1 stakes.

2.2 Network Structure

The classical notions of degree and clustering of nodes in a network need adaptation to properly describe a graph with weighted and directed edges.

2.2.1 Degree Distribution

Each node in our network has inward pointing edges and outward pointing edges. Each inward pointing edge represents money flowing inward, while an outward pointing edge represents money flowing outward. The total in degree of an edge is the sum of all the weights of inward pointing edges, and is equal to the gross amount of money won by that player. Likewise, the total out degree of an edge is the sum of all the weights of outward pointing edges, and is equal to the gross amount of money lost by that player. The net degree of an edge is total in degree minus total out degree, and represents a player's net winnings or losses. Losing players have a negative net degree. Right of origin is winning population, left is losing population. Each of the stakes looks to follow a power law distribution, both for the populations of winners and losers. Maximum likelihood estimation of the power law exponent yields an estimate of $\alpha = 1.3$ across all levels of stakes, for both winners and losers.

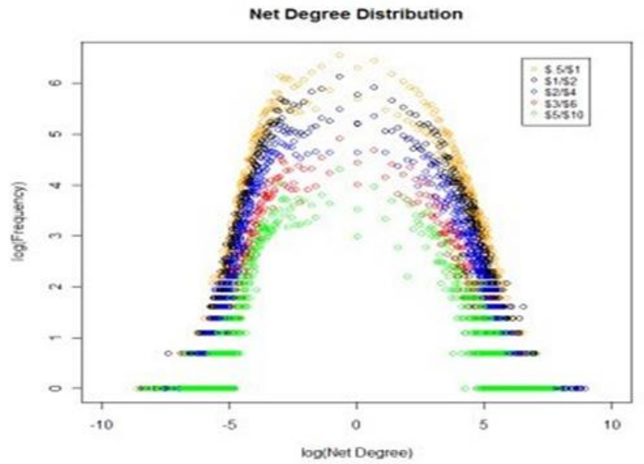


**Figure 1: Weighted degree distributions across stakes. Right of origin is winning population, left is losing population.**

2.2.2 Clustering

The clustering coefficient measures the tendency of nodes within a network to form local structures. The standard formulation of the clustering coefficient of a node takes into account only the existence of edges between nodes, but this does not provide a complete picture of the interactions in our network. Instead, we use a weighted generalization of the clustering coefficient proposed by Zhang et al (2005) [5][4]. Ci = P j,k wˆijwˆikwˆjk P j6=k wˆjkwˆik Ci is the clustering if node i, j and k are neighbors of i, and wˆi,j is the weight of the edge between nodes i and j, normalized by the weight of the largest edge in the network. This measure of clustering is higher for nodes in neighborhoods with very heavy weights, so nodes in communities that play many hands for large amounts have larger clustering coefficients. Plotting the clustering coefficient against net degree at different stakes, we can observe what types of players reside at those levels. Horizontal black line is average clustering, vertical black line is zero. Left of the vertical line are losing players, to the right are winning players. Figure 2 plots log(clustering) against net degree for each node at the $1/$2 stakes. Nodes in the top left quadrant are those with high clustering coefficient and negative net degree, i.e., players who play a lot, and consistently lose money. Note in the figure that there is a rather large community of dedicated losing players. Plots for each of these stakes reveal that $1/$2 games have by far the most “fish”, and as one moves up the ladder in stakes, the games become tougher, with fewer people consistently losing so much money. At $3/$6, only 3 players out of the player pool of 6,342 won or lost in excess of 2,000 big blinds. As shown in the plot above, a healthy population exists on both sides of this interval at $1/$2.

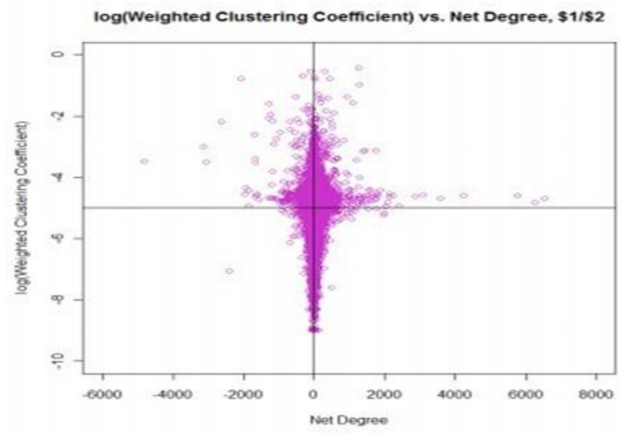


**Figure 2: Weighted clustering against net degree. Horizontal black line is average clustering, vertical black line is zero. Left of the vertical line are losing players, to the right are winning players.**

2.2.3 Pagerank

A node’s Pagerank score provides another measure of the influence or connectedness of a point. A node with a high Pagerank score has many incoming links from other nodes with high Pagerank scores, thus one could reasonably expect high Pagerank nodes to form communities of “regulars” who play many hands together. We hypothesize that these regulars are often superior in skill to casual players at their stakes, and a high Pagerank score should correlate posivitively with net degree. Plotting net degree against Pagerank score and fitting a simple least squares regression confirms this is the case. We observe this phenomenon at every level. Even at $3/$6 where the correlation is the weakest, we still observe βˆ1 = 69650, se(βˆ1) = 5190, with a t value of 13.48 on the regression coefficient for Pagerank score, indicating there is a definite positive correlation between Pagerank and net degree. This is evidence in support of our hypothesis that nodes with high Pagerank correspond to players with a higher degree of poker skill.

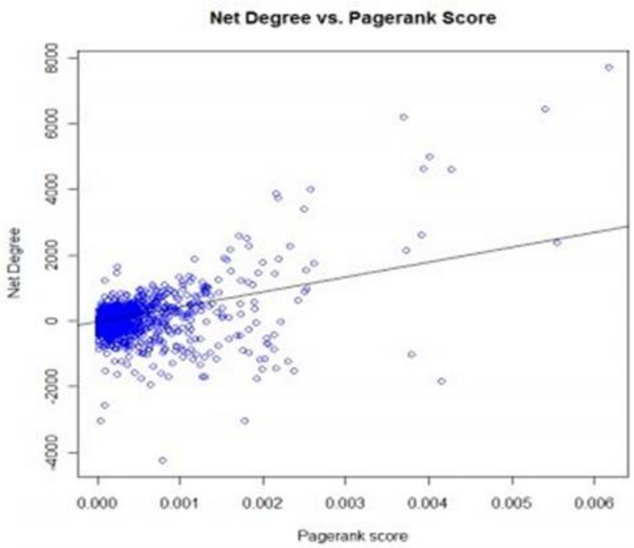


**Figure 3: Net degree versus Pagerank score for $2/$4 stakes. There is a clear positive correlation. For $2/$4, the regression coefficient on Pagerank score is statistically significant well beyond the .001 level. $\hat{\beta}_1 = 454901$, $se(\hat{\beta}_1) = 8299$, with a t value of 54.81**

  

## 2.3 Probabilistic Model for Network Evolution

We next sought to use the poker hand data to understand the microevolution of the poker social network. The goal was to find a generalizable probabalistic model with maximum likelihood of the time ordered hands. Previous work has been done applying preferential attachment models to the social networks of FLICKR (flickr.com), DELICIOUS (del.icio.us), YAHOO! ANSWERS (answers.yahoo.com), and LINKEDIN (linkedin.com) with success. However, in each of these cases, the creation of an edge comes at no explicit cost to either party. In a poker network, the interactions involve one player losing money to another, or winning from another, and thus the criterion for choosing who to interact with differ from the previously mentioned networks. We investigated ways of capturing this difference in the preferential attachment model. The model of preferential attachment differs slightly from those used in J. Leskovec, et. al, in that a new interaction could be between neighbors, rather than solely between nodes that do not already have an edge between them.

The model is as follows: node A is selected uniformly at random from the graph. With probability q, node B is chosen uniformly from the rest of the nodes in the total network. With probability $1 - q$, a neighbor of A is selected uniformly at random. With probability p, this neighbor is chosen as node B, and with probability $1 - p$, node B is chosen uniformly from the neighbors of the neighbor of A, excluding node A. The direction of the edge between nodes A and B is then chosen uniformly, with probabilities 1/2 for each direction. The model is summarized in figure 4.

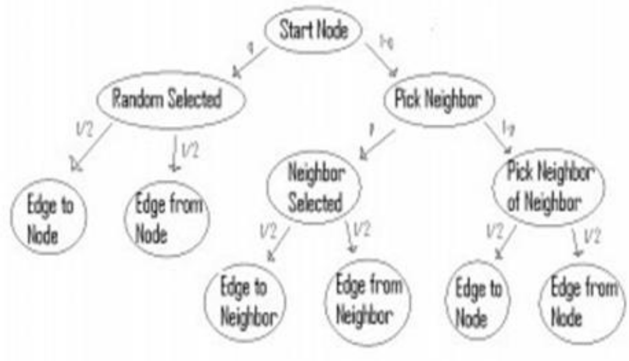


**Figure 4: Probability model for network evolution**

We then modified this preferential attachment model to no longer select the neighbors uniformly. Instead, the neighbors were weighted according to the net edge between node A and the neighbor. So if node A previously lost a large amount of money to neighbor C, then node A is unlikely to play against neighbor C again. $w(C) = \exp(Xab)$ Xab = net previous profit or loss of node A from node B These weights are used to form a new distribution for which the neighbor of node A is chosen. The same process is used when selecting the neighbor of neighbors, as the incentives are assumed to be transitive. If node A lost money to node B, and node B lost money to node C, then there is a disincentive for node A to play against node B. We found a much larger loglikelihood for the preferential attachment with uniform neighbor selection than for the modified, netprofit neighbor selection. This suggests that our weighting criterion is not a good choice for encapsulating previous experience. We then tried to improve the preferential attachment by predicting the edge directions between two nodes once they have been selected. We did this using the pagerank algorithm, generating the scores for nodes A and B . The pagerank algorithm was selected as regression models showed that produced scores were positive correlated with netprofit, the scores intuitively seemed to suggest skill levels, and the algorithm was quickly able to be prototyped and tested. We used the previously
described uniform neighbor preferential attachment model, but now used the scores to predict the edge direction instead of using uniform probability 1/2. The neighbor selection distribution was instead: p(edge B to A) = scoreA/(scoreA + scoreB) At each iteration, nodes A and B are selected according to the original preferential. This model produced a likelihood that was close, but still consistently worse than the unmodified preferential attachment model. These results held true at all levels of stakes. The experiments show that our probabilistic model of weighting neighbor selection by netprofit or weighting edge direction selection by Pagerank do not provide good results for the poker network evolution. Going forward, before performing more experiments it would be most useful to develop a metric that is better able to capture the desired edge prediction. Our likelihood evaluates which nodes create the edge and the edge’s direction, but does not predict the transaction’s magnitude. A probabilistic model that could accurately predict the magnitude of the win or loss would be preferred in a setting where a player may lose many hands, but win big on a few to still have netprofit. Once an agreed upon metric is found, then many other strategies can be tested for evaluating skill and predicting transactions. One possibility is to use logistic regression on the edges in the common neighborhood of node A and node B. Other methods for evaluating a total rank can also be tested, such that edges tend to point from lower to higher skill. These parameters could then be used to both predict edge direction, but also to understand how opponents are chosen.

## 3. CONCLUSIONS

Our work in exploring the world of online poker gave us some interesting insight into the network structures that form within the community. Although our attempts to model the microevolution of the network did not produce ideal results, they still left us with a better understanding of the problem, and ideas for exploring this topic further.

  