Statistical Network Analysis of a Real-time Strategy Multiplayer Online Battle Arena Game

Kanghong Shao

Department of Statistics, University of California, Los Angeles

Abstract

In recent years, Multiplayer online battle arena (MOBA) games have become one of the most popular and widespread online video games over the world [1]. It provides large amount of interesting relational data for analysis. In this project, I conduct various statistical network analysis of the relational data among heroes for Dota 2 (a MOBA game). The descriptive analysis of centralization shows high centralization in the network. ERGM models with covariates suggest that forming of ties depends highly on a character's KDA ratio performance over another. Latent position models with edge covariate indicates that people tend to play stronger characters on the peripheral area of the network more than the weak ones near the center. Latent position clustering models provides latent positions, which has a core-and-peripheries shape and a "layered" clustering pattern. Additionally, it is found that the network has a preference for intransitive ties. Finally, several issues raised by the results are discussed and suggestions are provided for future improvement and analysis.

Several issues are raised by the results through the analysis. Firstly, KDA ratio seems to be the only important covariate that impacts the pattern of ties, but it is a relatively naive statistics developed by non-statistician, which has many drawbacks and limitations in capturing the overall impact of a hero in the game. It is very "stats-driven" and many aspects of the game can not be reflected by this simple ratio. One possible future solution is to collaborate with experts from gaming industry and develop better statistics/measurements to evaluate hero's performances. Secondly, win rates might not be the best measurement to represent a hero's advantage again another. Finally, when interpreting the "layered" clustering pattern, I make a possible theory that there is a underlying force depending on the powerfulness of actors. If one actor powers over the other actor, the force draws them together. If two actors have similar level of powers, the force pushes them away. For further analysis, better measurements for the overall impact of characters and a hero's advantage again another can be studied to improve the network analysis. Future studies to better understand the "layered" clustering pattern, and the results can be applied to fields other than gaming.

1 Introduction

In today's world, digital life has become one of the most important parts of many people's life, especially the youth and young adults. It takes various forms, such as social media, online shopping, and website browsing. Many studies and researches have been done for them, but there is one form which has not been well studied yet, online video gaming. Online video gaming, though appears to be less essential and common than the others, might actually be the one that involves most interesting intercalations between individuals. Therefore, it is a great source for statistical network analysis of relational data, and may provide important insights about how people interact with the each other through the gameplay mechanics.

Multiplayer online battle arena (MOBA), also known as action real-time strategy (ARTS), is a sub-genre of strategy video games in which a player controls a single character in a team who compete versus another team of players [2]. In recent years, with the development of eSports, MOBA games have become one of the most popular, widespread and active competitive online video games over the world with the huge successes of DotA, Dota 2, League of Legends (LoL) and Heroes of the Storm (HotS) [1]. Millions of people over the world play them on a daily basis in public games, and hundreds of teams from across the world playing professionally in various leagues and tournaments for millions of U.S. dollars as prizes.

In this project, I conduct statistical network analysis of one of the most popular and well-developed MOBA game, Dota 2. It is played in matches between two teams of five players, with each team occupying and defending their own separate base on the map. Each of the ten players independently controls a different game character, called a "hero". All the heroes have unique abilities and differing styles of play. In a Dota 2 game, players collect gold, earn experience points and build items for their characters in order to successfully battle the opposing team's heroes, who are attempting to do the same against them. A team wins the game by being the first to destroy a large structure located in the opposing team's base, called the "Ancient". Figure 1 below shows the simplified typical map of a MOBA genre game. Yellow lines are the "lanes" where action is focused; blue and red dots are the defensive "towers/turrets" that defend them; light-colored quarter circles are the teams' bases; blue and red corners are the structures whose destruction claims victory [2]; green areas are jungle areas where neutral creeps live and players can kill them to earn gold and experience points. Despite MOBA games? high popularity, great impact and numerous amount of available matches data and network data, very little statistical analysis has conducted about them. This project carries out statistical analysis of the relational data about player-controlled characters, and attempts to fill in the gap between MOBA games and statistical study.

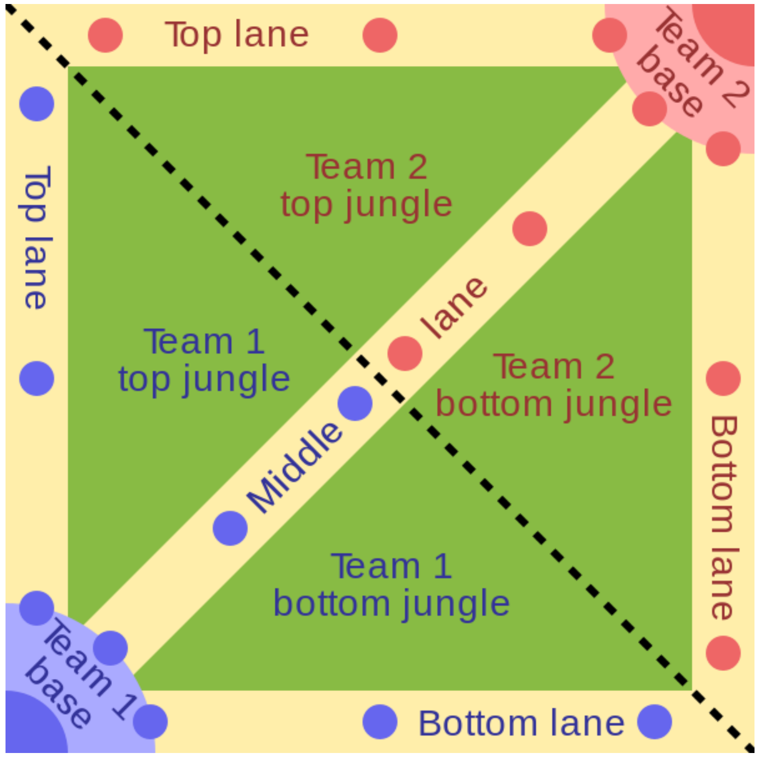

Figure 1: Typical map of a MOBA genre game.

The project address two main issues. The first goal is to find network patterns and structures for the relations between different game characters which players control in public games. And the findings can be useful in making future improvement to the game create a more balanced underlying mechanics to improve player?s game experiences. The second goal is to get possible insights about the player interaction forces underlying the character interaction in the games. For future study, a more comprehensive analysis of the direct player network can be done to look more closely about player interaction, which the project does not include due to limited time and computing capacity.

1.1 Data and Processing

The original data is collected from OpenDota.com and DOTABUFF.com [3] [4]. The dataset for analysis consists of two components. First component is an edge list of all the matchup records between each character and another. Since the number of characters is 115, there are 115 $\times$114 = 13110 pairs in total. Denote all the characters as i for i from 1 to 115 against all other characters j for $j \neq i$, and a pair of characters matchup record as $\{i, j\}$. Also denote the win rate of a primary character i against an opposing character j as $y_{i,j}$. Notice that the win rate $y_{i,j}$ of i against j is equal to $1 - y_{j,i}$ (j against i), therefore I will only keep the matchup record $\{i, j\}$ since the reverse record $\{j, i\}$ is known automatically.

After carefully examining the matchup features available from the data, the win rate $y_{i,j}$ is used as the outcome variable to evaluate the relations between character i against j. Instead of conventional relation in network analysis such as friendship, protein binds, communication, sexual contact, proximity, migration rate, alliance/conflict and so, win rate captures a character?s overall performance is again another hero when played by numerous people across public games. In other words, how much a character ?powers? another hero. Since $y_{i,j}$ of i against j is equal to $1 - y_{j,i}$ (j against i), matchup records with win rates $y_{i,j} < 0.5$ are excluded from the dataset. The theoretical number of matchup pairs should be $13110/2 = 6555$, but since there are some hero matchups where the $y_{i,j}$ are exactly 0.5, they are also excluded from the dataset, leaving number of matchup pairs is 6547.

1.1.1 Construct Edges and Edge Covariate

The network is still too dense since there are 6547 ties between 115 nodes. For simplicity, for the most models and analysis, win rates $y_{i,j}$ are considered as binary variable with values of $\{0, 1\}$. Therefore, a cut-off value of win rate is required to transform the continuous number into $\{0, 1\}$. To achieve this, the quantile table of the win rates is examined to determines the cut-off win rate, as shown in Table 1. A cut-off win rate τ is chosen so that $y^*_{ij} = 1$ if $y_{ij}\ \tau$. The choice of τ is completely empirical. For the purpose of this project, 0.6 is chosen and the corresponding remaining quantile is 6.42%. Theoretically, any reasonable choice of win rate is fine, but choosing a lower win rate results in an extremely dense network where there are too many edges for only 115 nodes for efficient analysis and visualization. The final edge list is of length 420 and the corresponding sociomatrix A is created as a 115×115 matrix with 420 non-zero entries. In addition, an edge covariate "diff KDA" is assigned to each each, which is equal to the different between the primary character i and the opposing character j. The notion of KDA will be explained below.

Quantile	50%	55%	60%	65%	70%	75%
Win Rate	0.53620	0.54110	0.54600	0.55070	0.55630	0.56285
Quantile	80%	85%	90%	95%	100%	
Win Rate	0.56930	0.57750	0.58840	0.60618	0.68820	

Table 1: Quantile Table of Win Rates

1.1.2 Construct Nodal Covariates

The second component is the nodal covariates for each character. The original nodal dataset has 32 covariates. Based on knowledge of the game and after combining multiple features into new variables, there remains in total, namely "Position", "KDA Ratio", "KDA Group", "GPM", "XPM", "Hero DamagePM" , "Tower DamagePM", "Hero HealingPM". "Position" means which position or role a character usually plays by players. There are 6 empirical catogaries: carry, mid laner, off laner, jungler, supporter and mix. Carries have the highest priority for levels (experience points) and gold on the team. They are usually the weakest on the team before they have an item or two. They often have abilities that let them scale well with bonus stats gained from items, allowing them to deal a ton of damage later in the game. Mid laners have the second most need for gold and levels. A position 2 hero will generally go to mid lane and use their high gold and experience income to control the pace of the game. Off laners generally builds toward team fighting items or initiation items. They are often placed in the off lane where the main goal is to gain is experience points and increase levels. Junglers are usually the farming support on the team. They usually gain their gold and levels by farming in the jungle. They build towards support items that benefit the team. The Junglers also help with pure support items when the support cannot afford to. Supports use gold on mainly pure support items. As a hard support this sacrifice benefits the rest of the team, helping them excel [5]. "KDA Ratio" is the average KDA ratio of a character across games. The formula to compute it is $(K+A)/D$ where K is the number of kills (of heroes in the opposing team), A is the number of assists (for kills) and D is the number of death in a game. It is the most commonly used statistics to measure the performance of a player in a match, so the average KDA of a character by numerous players measures the average performance of a character. Furthermore, "KDA Group" is constructed to categories heroes into 3 groups: low KDA, mid KDA and high KDA. The corresponding thresholds are 2.35 (33.3%) and 2.64 (66.7%), which are chosen to evenly separate the total number into three groups. GPM means Gold/Minute, and XPM is Experience/Minute. Hero DamagePM , Tower DamagePM and Hero HealingPM stand for Hero Damage/Minute, Tower Damage/Minute and Hero Healing/Minute, respectively.

2 Analysis and Results

2.1 Summaries of the Networks

2.1.1 Visual Summaries

First of all, the network is visualized into a 2-D plot as shown in Figure 2. We observe that there is one giant component containing almost all the nodes (characters) and 3 isolated nodes on the side, which are ?Death Prophet?, ?Keeper of the Light? and ?Queen of Pain?. The 3 isolates indicate that they have some ?safe? positions when playing against any other characters. In other words, they don?t really power over any other characters, but also can not be well countered by any others either. In addition, other characters have the least influence on them, so as long as no updates are made on them, they will likely to remain similar status when other characters are buffed or nerfed by new game updates.

The largest component of the network (see Figure 3) consists of the rest 112 nodes, and has some very interesting patterns. Firstly, there appears to exist a sphere shape or high-degree star shape in the right part of the network, which has two nodes at center position that both have extremely high in-degrees pointing towards them from many surrounding nodes. (Notice that the exact position of the network in visualization plot is at random, so the nodes does not have to be on the right part.) This means that players who controlled them are generally overpowered by many others, represented as peripheral nodes, cross public DOTA2 games on the internet. Secondly, there seems to exist a large cloud of nodes in the left part aside the sphere. It has more ties and more complex patterns of ties among the nodes. Many nodes also form ties with the nodes of the sphere/star. In addition, there are several nodes including "meepo", "dark-seer", "pangolier", "slark" and perhaps "ember-spirit" which are relatively far away of either the cloud and the sphere. They all have only in-degree ties pointing towards them. This indicates that they are considered as "weaker" character across public games. So far, the original plots of network show some interesting patterns in it. Further analysis are carried out to dig deeper into them.

One of the core issue when investigating a network is: how do individual node's position varies? One manner in which they vary is the extent to which they are "central" in the network. The out-degree distribution is a numerical count of numbers of out degree ties for each actor/node, and similarly for in-degree distribution. They summarize the densities of ties of the population of nodes. It providers with information about the overall activity or extent of involvement for individual actor in relation to others, and how central the individual node is in the entire network. In particular, for the DOTA2 characters, the out degree reflects how strong or powerful a player's game character is, while the in degrees measures how weak or powerful they are. The plot of the distributions are shown in Figure 4. The concept "central" is not of the same conventional meaning in this extent. Firstly, it can be interpreted as how "unbalanced" the individual hero is, whether they are too powerful or too weak in the game. Secondly, it also illustrates how influential it is if its abilities and properties are changed in the game mechanics, but also may be subject to a great deal of influence/changes from other characters. We discovered that while there are several different characters who have relatively high out degrees, only two nodes have outstandingly high in degrees, whose corresponding characters in the game are "IO" and "Lone Druid". Their in-degrees are significantly higher than any another characters, which suggests that players? experiences when playing these characters are much worse in public games, and that the game developer and designer should probably make large changes to buff them. On the other hand, similar logic should also be applied to game characters whose power over other ones is too much, in which cases they should be reasonably nerfed. Notice that there it is not saying that a static balance should be the goal, but a dynamic equilibrium of the network is wanted.

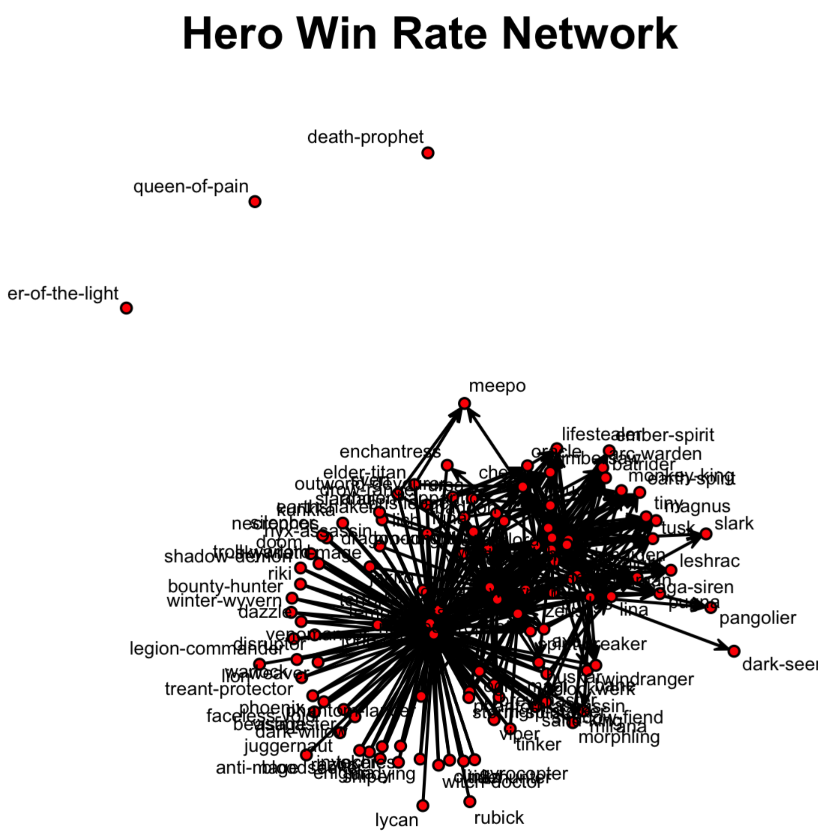

Figure 2: Network of Hero Win Rates

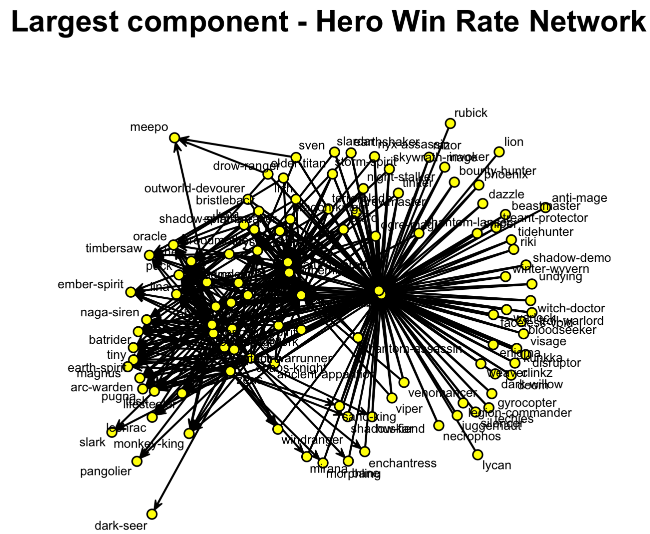

Figure 3: Largest Component in Network of Hero Win Rates

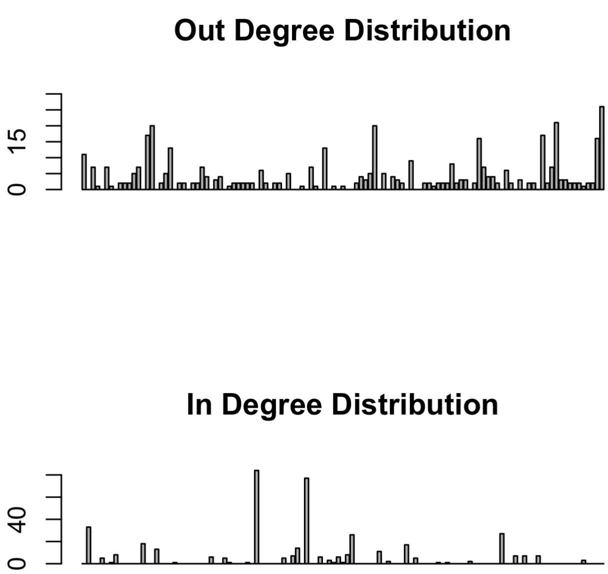

Figure 4: Out and In Degree Distributions

2.1.2 Numerical Summaries

The entire network data is a directed network with 115 vertices. The number of possible ties in the network is 13110, and the number of actual ties is 420. The density of ties is equal to 0.032. The data is based on all the public matches from November 17, 2017 to December 13, 2017. This time range is selected because there was an important game version 7.07 released on October 31, 2017 and the most recent updates 7.07c was on November 17, 2017. So the dataset only contains public games after the most recent game update. The minimums of in and out degrees are 0, and the mean of in and out degrees is 3.65. The maximum of in degree is 84, while the maximum of out degree is 26.

2.2 Patterns of Centralization of the Actors

Besides degree counts of direct ties, there are several other measurements for actor centrality, namely betweenness, closeness and eigenvector centrality. Betweenness measures tendency of ego (the focusing actor) to reside on shortest paths between third parties; closeness measures how close is an actor to other ones, measured by geodesic distances; and eigenvector centrality is defined such that the centrality of each actor is equal to the sum of the centralities of its neighbors, attenuated by a scaling constant. The distributions of betweenness, closeness and eigenvalue centrality are in alignment with the degree distribution, shown in Figure 5, Figure 6 and Figure 7. Although they are not measuring the exact same thing, all the them show that the two actor "IO" and "Lone Druid", as illustrated by the degree distribution, are the most central nodes of the network.

Based the node-level centralities, network centralization is computed as the network-level measurement of how centralized the entire network is. Table 2 shows the summary of 4 types of centralization of the network. To summarize the results, the character network has high centralization in general. This is reflected by both the visualization and different types of centralization measurement. Indeed, although the "cloud" component is not clear yet, about half of the network looks visually centralized at the two nodes in the center.

Table 2: Network Centralization

Measurement	Degree	Betweenness	Closeness	Eigenvalue
Centralization	0.7017199	0.4191147	0.704382	0.4224704

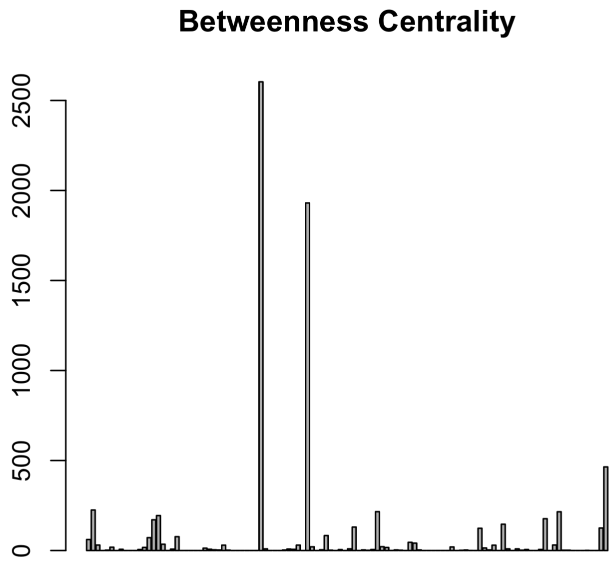

Figure 5: Betweenness Centrality

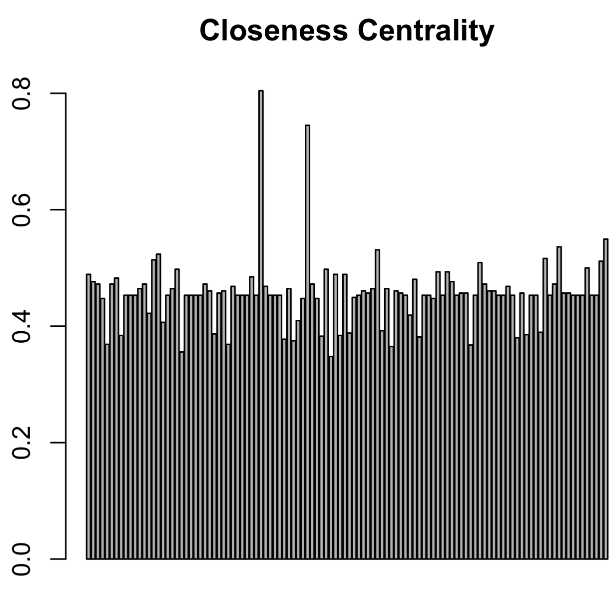

Figure 6: Closeness Centrality

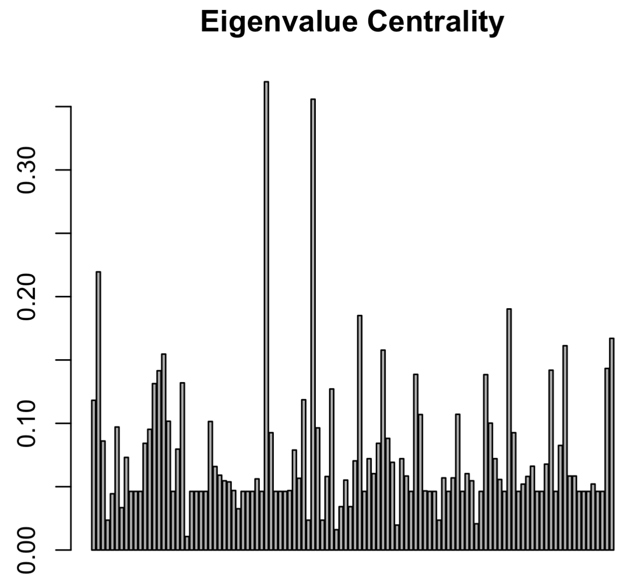

Figure 7: Eigenvalue Centrality

2.3 Impact of Covariates on Patterns of Ties

2.3.1 Impact of Nodal Covariates

The character network is different from a conventional friendship or other positive relationship networks in which actors with similar properties are more likely to form ties and tend to be close to each other. Instead, powerful actors have high propensity to form ties with weaker actors and are less likely to form ties with other actors with similar level of powerfulness. Therefore, a plausible underlying assumption to make for the character network is that a covariate x_i which reflects how strong or weak a character is played across games in general would have a strong impact on the patterns of ties. More precisely, a comparison between actor i's and j's powerfulness level x_i and x_j, by assumption, should influence the probability to form a directed tie from i to j.

To inspect the impact of nodal covariates on patterns of ties of the network, exponential-family random graph (ERGM) models are used with chosen statistics. The basic assumption of ERGM models is that the structure in an observed graph y can be explained by any statistics $g_k(y)$ depending on the observed network and nodal attributes [6]. Therefore, probability distribution of the set of possible graphs is:

$$P(Y = y) = \frac{exp\{\sum_{k=1}^{K} \theta_k g_k(y)\}}{c(\theta)} \tag{1}$$

where $\theta_{1,2,\ldots,k}$ are parameters, $g_{1,2,\ldots,k}(y)$ are statistics, and $c(\theta)$ is a normalizing constant. These models represent a probability distribution on each possible network on n nodes. The Intuition is that the ERGM model places more or less weight on graphs with certain features, as determined by θ and g [6].

The pre-constructed categorical nodal covariates "KDA Group" and "Position" are used as nodal covariate x_i for fitting ERGM models. After fitting ERGM models with different combinations of the covariates, it turns out that a simple ERGM with nodal covariate "KDA Group". The result is shown in Table 3:

	Model 1
edges	-4.10^{***}
	(0.19)
mix.KDA_group.high KDA.high KDA	-0.31
	(0.32)
mix.KDA_group.low KDA.high KDA	-1.99^{**}
	(0.61)
mix.KDA_group.mid KDA.high KDA	-0.94^{*}
	(0.37)
mix.KDA_group.high KDA.low KDA	1.94^{***}
	(0.21)
mix.KDA_group.low KDA.low KDA	1.09^{***}
	(0.23)
mix.KDA_group.mid KDA.low KDA	1.55^{***}
	(0.21)
mix.KDA_group.high KDA.mid KDA	0.66^{**}
	(0.24)
mix.KDA_group.low KDA.mid KDA	-1.02^{**}
	(0.38)
mix.KDA_group.mid KDA.mid KDA	
AIC	3328.05
BIC	3402.86
Log Likelihood	-1654.03

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: ERGM Model Fit with Covariate: KDA Group

The coefficient estimates show several patterns of forming ties. The coefficient estimate of edges term is -4.10 with significance, so the corresponding probability of forming a tie between two random nodes is small, which means that the overall probability of a tie forming between two random nodes is low. The coefficient estimates of high KDA vs. mid KDA, high KDA vs. low KDA and mid KDA vs. low KDA are all positive values with significance, whereas the inverse pairs all are negative values with significance. This means that ties are more likely to occur based on KDA outperforming. Moreover, there is a general pattern for ties to form more likely if the primary character's KDA outperform more than the opposing character. e.g. high KDA vs. low KDA (1.94) has higher likelihood than mid KDA vs. low KDA (1.55).

In addition, the estimate of low KDA vs. low KDA is 1.09 with significance. This indicates that characters in low KDA group are more likely to form ties between each other. Notice that the coefficient estimates of the high KDA vs. high KDA is negative but with non-significance p-value, which means that there is no particular positive nor negative tendency of forming times between high-KDA character. Finally, estimate of mid KDA vs. mid KDA is NA, meaning there is not such ties.

2.4 Latent Positions of Network

2.4.1 Latent Position Model

Let $\{z_i\}$ be the positions of the actors in the social space R^k; $\{x_{i,j}\}$ denote observed characteristics that may be dyad-specfic and vector-valued. For simplicity and interpretability, $d_{i,j} = |z_i - z_j|$, where the z_i's are positions of the nodes in R^k and the distance is Euclidean. Then the probability of a tie follows a logistic model [7]:

$$logodds(Y_{i,j} = 1|z_i, z_j, x_{i,j}, \beta) = \beta_0^T x_{i,j} - \beta_1 |z_i - z_j| \tag{2}$$

Model-based estimation of latent positions for networks provides a formal network model that incorporates the notion of positions. It reduces the feature space into a lower dimension such as 2D or 3D and gives the estimation of positions of all the characters played by players across public games, which can be thought as the underlying characteristics positions for characters in the game. Table 4 and Table 5 show that latent space model in 2D space has a better BIC than that in 3D space. Figure 8 is the plot of latent positions of each character from the fitted simple latent space model with euclidean distances in 2 dimensional space.

Table 4: Summary of Latent Position Model Fit in 2D Space

	$Estimate$	2.5%	97.5%	$2 * min(Pr(>0), Pr(<0))$
(Intercept)	-1.3325	-1.5358	-1.1277	<2.2e-16 ***
Overall BIC	3856.015			
Likelihood BIC	3317.806			
Latent space/clustering BIC	538.2086			

Signif. codes: 0 $***$ 0.001 $**$ 0.01 $*$ 0.05 . 0.1 1

Table 5: Summary of Latent Position Model Fit in 3D Space

	Estimate	2.5%	97.5%	$2 * min(Pr(>0), Pr(<0))$
(Intercept)	-0.87088	-1.12769	-0.6088	<2.2e-16 ***
Overall BIC	4060.505			
Likelihood BIC	3258.799			
Latent space/clustering BIC	801.7062			

Signif. codes: 0 $***$ 0.001 $**$ 0.01 $*$ 0.05 . 0.1 1

Compared to the original plot in Figure 1, the fitted positions of the network has a clearer pattern of centralization with the center being nodes which has high number of in degrees. The network also seems to have a more complete periphery, most of which are nodes that send ties towards the center nodes. In addition, the large cloud of nodes in the left part aside the sphere that appears in the original network plot does not exist. Notice that there does still appears to a small chunk of nodes at the bottom-right corner that is very close to each other and to the center nodes. It is actually corresponds to the large cloud in the original plot after matching labels. In the plot, different colors of the nodes show the different positions/roles each character plays most frequently in the games. It is very hard to visually identify any clear patterns about which kind of characters are more likely to appear in the center or in the periphery. Furthermore, notice that although no clear clusters can be observed, the network plot seems to exist, to some extent, a central core and layers of periphery.

In order to illustrate the relationship between where a character' position is in the latent space and players' preference for choosing that character, the vertex sizes are adjusted proportional to the number of games in which the character is played. The plot shows that besides the largest blue nodes near the center (which is a mix-type character "Pudge"), most of the larger nodes lie in the peripheral area of the network. This alignments with the intuition that people tend to play characters who are stronger and have power over other players.

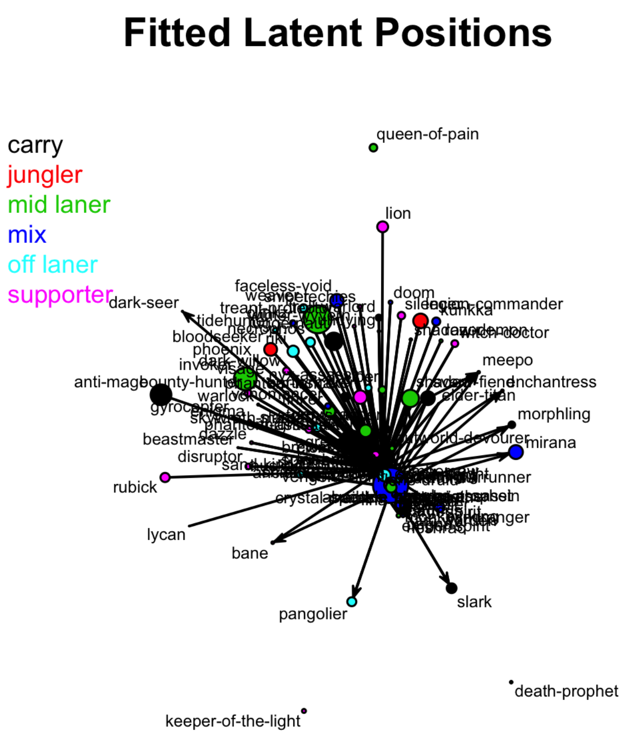

Figure 8: Plot of Fitted Latent Positions

2.4.2 Impact of Edge Covariate in Latent Position Model

To further account for edge covariate impact on patterns of ties, an additional edge attribute ”diff KDA” is added to original latent position model. The summary of the fit is shown in Table 6. The overall BIC has been significantly improved by including the edge covariate ”diff KDA”.

Table 6: Summary of Latent Position Model in 2D with Edge Covariate: diff KDA

	Estimate	2.5%	97.5%	$2 * min(Pr(>0), Pr(<0))$
(Intercept)	-1.3325	-1.5358	-1.1277	<2.2e-16 ***
diff KDA	18.3576	16.5256	20.2758	< 2.2e-16 ***
Overall BIC	2150.813			
Likelihood BIC	1253.371			
Latent space/clustering BIC	897.442			

Signif. codes: 0 $***$ 0.001 $**$ 0.01 $*$ 0.05 . 0.1 1

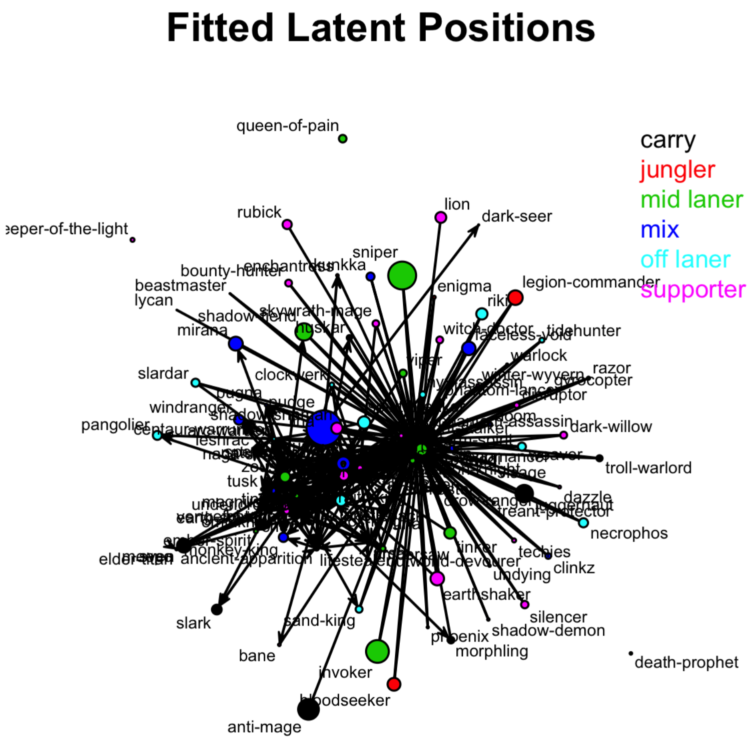

Figure 9: Plot of Fitted Latent Positions with Edge Covariate: diff KDA

Compared to that of the simple latent position model, the fitted positions of the latent position model with edge covariate captures both the sphere pattern and the "cloud" pattern saw in the original plot (Figure 9). In addition, It has a pattern of centralization with seemingly 2 cores instead of 1. The plot also shows that besides the largest blue nodes near the center, most of the larger nodes lie in the peripheral area of the network.

To sum up, both latent space models indicate that people tend to play characters on peripheral area of the network, who are in general stronger than those near the center, and hence they would have more power over many other players. However, although it seems to be a pattern of centralization with 2 centers, it is still not clear what the structure or clustering pattern it has.

2.5 Network Clustering

The character network is different from a conventional friendship or other positive relationship networks in which actors with similar properties are more likely to form ties and tend to be close to each other. Instead, powerful actors have high propensity to form ties with weaker actors and are less likely to form ties with other actors with similar level of powerfulness.

Latent position cluster models are fitted to the character network in order to further analysis possible clustering patterns in it. The models model the latent positions as clustered into G groups [8]:

$$z_i \sim \sum_{g=1}^{G} \lambda_g MVN_d(\mu_g, \sigma^2 I_d) \tag{3}$$

The advantage of this model compared to latent position model is that it not only captures position, transitivity, homophily on attributes, but also clustering as well. Bayesian model selection (approximated by a version of BIC) can be used to determines the number of groups. Figure 10 shows the BICs of latent position cluster model with 2, 3, 4 and 5 groups, with 4-group model having the lowest BIC. The corresponding fitted latent positions of the network in plotted in Figure 11. The results of its Goodness-Of-Fit diagnostics are shown in Figure 12 and Figure 13, which suggest that the fitted model converges well.

Because the fitted areas of clusters by the model has a pattern of "circles within circles", the conventional notion of clustering might not be appropriate for this particular network data. Instead, it offers evidence that the network has 3 to 4 "layered" clusters. To sum up, the character network has a pattern of "layered" clustering.

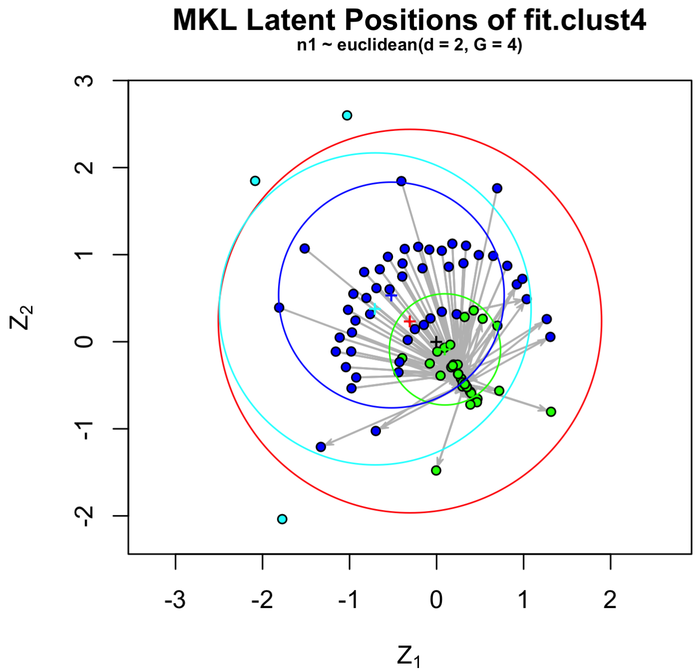

Figure 11: Plot of Fitted Latent Position Cluster Model of 4 Groups

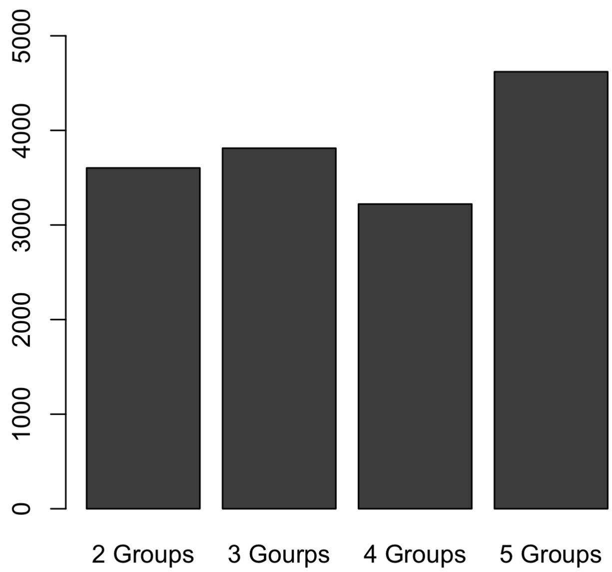

Figure 10: BICs over Different Number of Clusters

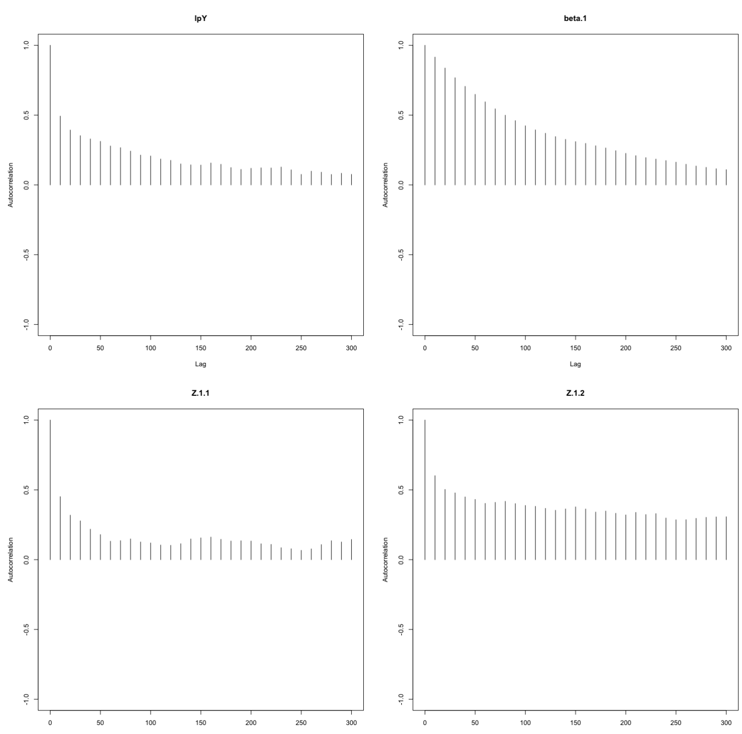

Figure 12: Diagnostics Latent Position Cluster Model of 4 Groups

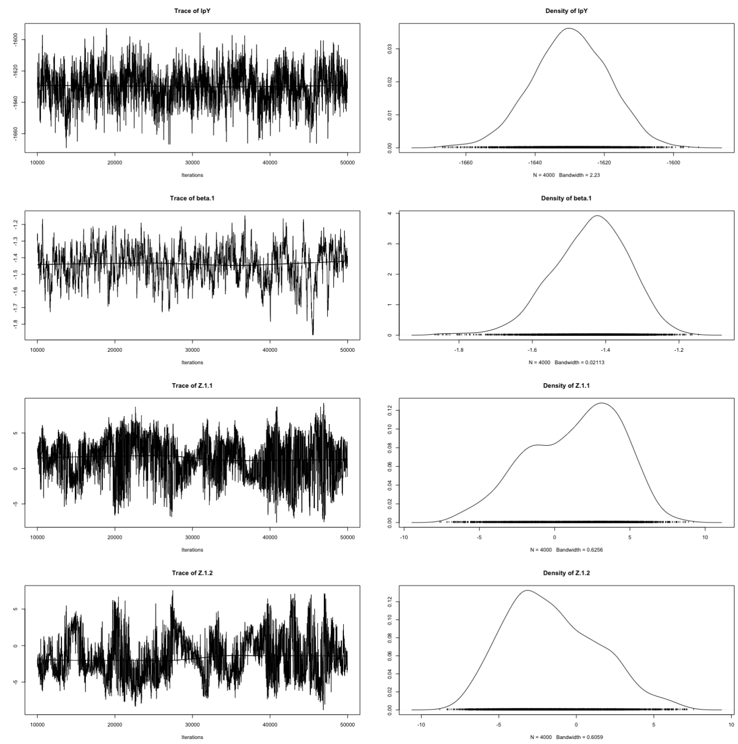

Figure 13: Diagnostics Latent Position Cluster Model of 4 Groups

To sum up, the result of the latent position cluster model shows that the clustering pattern of the character network is different from a convention network such as friendship. Because in powerful characters have high propensity to form ties with weaker actors and are less likely to form ties with other actors with similar level of powerfulness. To better understand the clustering, a possible assumption is that there is a underlying force which depends on the powerfulness of the actor in the network. If one actor powers over the other actor, the force draws them together. If two actors have similar level of powers, the force pushes them away. As a result, the weaker actors in the center area are drawing stronger actors towards them, but the stronger actors themselves are pushing away from each others. And hence, the network has a core-and-peripheries shape and "layered" clustering pattern.

2.6 Network Transitivity and Balance

The concept of statistical degree of balance in directed relations is applied naturally in the character network, that is: actors seek out transitive relations, and avoid intransitive relations. Assuming hero i powers over hero j and hero j powers over hero k, then hero i might also power over hero k.

Firstly, a frequency table of triad census is computed as the counts of triads by class in each one of 16 different categories of directed triads. The counts are shown in Table 7. Among all the 16 types of directed triads, only 003, 012, 021D, 021U and 030T exist in the network, with most common triads (besides 003 because it has no ties) being 012 and 021U.

Triad	003	012	102	021D	021U	021C	111D	111U
Counts	209831	26,701	0	2,011	8,349	0	0	0
Triad	030T	030C	201	120D	120U	120C	210	300
Counts	13	0	0	0	0	0	0	0

Table 7: Frequency Table of Triad Census

To inspect the balance and transitivity in the patterns of ties, a triad census model is fitted to the data. The results for coefficients are shown in Table 8. The coefficient estimates of vacuous triads 012, 021D, 021U are 0.07, -0.31 and 0.17. The coefficient estimate of transitive triad 030T is -0.43, which is negative with high statistical significance. Hence there are less transitive triads in this network than we would expect had we fixed the other sufficient statistics and randomly placed ties amongst the nodes. The change in log-odds attributed to transitive triads when we add an extra to a given edge is equal to -0.43 $\times$ number of transitive triads created. This contradicts to the intuitive assumption that if hero i powers over hero j and hero j powers over hero k, then hero i might also power over hero k. In fact, the overall triad census pattern means that there are less transitive triads than expected, which indicates intransitivity. So there is a preference for intransitive ties in the network.

	Model 1
triadcensus.012	0.07^{***}
	(0.00)
triadcensus.021D	-0.31^{***}
	(0.01)
triadcensus.021U	0.17^{***}
	(0.00)
triadcensus.030T	-0.43^{***}
	(0.01)
AIC	14617.97
BIC	14647.89
Log Likelihood	-7304.98

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 8: Statistical models

3 Discussion

In this project, I have conducted multiple statistical network analysis of the relational data (win rates) among heroes for one of a popular MOBA game, Dota 2. I first perform descriptive analysis to show the patterns of centralization. Result show that the network has high centralization in general and suggests further analysis about the network structure. By fitting ERGM models with covariates, I have found that forming of ties depends highly on a character's KDA ratio performance over another. It is more likely for ties to form if the primary character's KDA ratio outperforms the opposing character. Then I fit latent position models and explore the influence of edge covariate in these models. Results of model fitting suggest that including edge covariate of difference in KDA ratio significantly improves the overall BIC. On the other hand, the plots indicate that people tend to play stronger characters on the peripheral area of the network more than the weak ones near the center. To further analysis the network structure and clustering, I fit latent position clustering models and find that best model has 4 groups. The clustering pattern of the network is different from that of a convention network such as friendship and the fitted positions have a core-and-peripheries shape and a "layered" clustering pattern. Finally, network balance and transitivity are inspected and I find that contradicts to the intuition of transitivity relations, the network has a preference for intransitive ties.

Several issues are raised by the results through the analysis. Firstly, KDA ratio seems to be the only important covariate that impacts the pattern of ties, but it is a relatively naive statistics developed by non-statistician, which has many drawbacks and limitations in capturing the overall impact of a hero in the game. It is very "stats-driven" and many aspects of the game can not be reflected by this simple ratio. One possible future solution is to collaborate with experts from gaming industry and develop better statistics/measurements to evaluate hero's performances. Secondly, win rates might not be the best measurement to represent a hero's advantage again another. Finally, when interpreting the "layered" clustering pattern, I make a possible theory that there is a underlying force depending on the powerfulness of actors. If one actor powers over the other actor, the force draws them together. If two actors have similar level of powers, the force pushes them away. For further analysis, better measurements for the overall impact of characters and a hero's advantage again another can be studied to improve the network analysis. Future studies to better understand the "layered" clustering pattern, and the results can be applied to fields other than gaming.

References

[1] Park, Hyunsoo, and Kyung-Joong Kim. "Social network analysis of high-level players in multiplayer online battle arena game." International Conference on Social Informatics. Springer, Cham, 2014.

[2] Wikipedia contributors. "Multiplayer online battle arena." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 1 Dec. 2017. Web. 16 Dec. 2017.

[3] Opendota. http://www.opendota.com.

[4] Dotabuff. http://www.dotabuff.com.

[5] Contributors. "Farm Dependency." Liquipedia. N.p., 21 Dec. 2016. Web. 15 Dec. 2017.

[6] Mark S. Handcock. "Introduction to Exponential-family Random Graph Models (ERGMs)." STATS218: Statistical Analysis of Networks. University of California - Los Angeles, Sep. 28, 2017.

[7] Mark S. Handcock. "Latent Position Models for Networks." STATS218: Statistical Analysis of Networks. University of California - Los Angeles, Nov. 30, 2017.

[8] Mark S. Handcock. "More on Latent Position Models." STATS218: Statistical Analysis of Networks. University of California - Los Angeles, Nov. 30, 2017.