The Contagion of Cheating in Online Gaming

A dissertation submitted by candidate 12357

for the MSc in Applied Social Data Science

Department of Methodology

London School of Economics and Political Science

2019

Candidate number: 12357

Supervisor: Milena Tsvetkova

Word count: 8,743

08.08.2019

Abstract

Behavioural contagion is a domain that social scientists have long explored as the pivot of population processes within social systems. Considering the increased interest in the topic of antisocial behaviour in the online space, this study examines how antisocial behaviour spreads online, particularly in the online gaming world. Among various types of antisocial behaviour in online gaming, this study focuses on cheating. While relatively more is known empirically about how online game cheating spreads among friends, much less is known about the contagion of cheating among strangers. With a dataset consisting of 1,179,537 matches, this study uses networks to understand how cheating spreads among players in the game world of *PlayerUnknown's Battlegrounds*, which is an online first-person shooter game. In an effort to overcome the presence of missing information, this paper attempts to estimate the time of cheating adoption before the data analysis. Focusing on the generalised reciprocity mechanism, this study undertakes a network analysis to examine how cheating spreads through contacts with strangers who are randomly matched as competitors. To build a large-scale temporal network, pairs of players are connected by directed time-stamped edges if one player kills the other. The extent of damage caused by cheating to victims is quantified and added to the analysis. Evidence suggests that victims of cheating are likely to adopt cheating after suffering severe loss and damage by cheaters. The paper includes implications of the results and makes recommendations for further research work. The importance and originality of this study are that it explores the contagion of online game cheating through contacts with strangers, which has not been closely examined, and shows that the severity of the damage caused by cheating is an important factor that triggers the spread of cheating.

1 Introduction

The main challenge faced by the online gaming industry is the prevalence of cheating. Many cheating methods have been used since the first phase of the popularisation of computer gaming when most players played single-player games. At that time, cheating was not a serious issue because players who cheated did so against computer opponents in single-player games. However, since online multiplayer games have gained massive popularity, cheating has become a major problem because it harms the game experience and enjoyment of other players. With an increased demand for tools and methods to support cheating, online game cheating today involves a lucrative but illegal cheat code development business. Cheat code providers make big profits from selling cheat codes to players through their websites (Maiberg, 2014), while some players download free cheating tools online even at the risk of infection from computer viruses (Sampson, 2018).

In most games, players become skilful by playing many matches and learning from experience. However, players who use cheating tools can unfairly enhance their abilities and thus increase their chances of winning matches without devoting the necessary time and effort. Many types of cheating tools exist to assist cheaters to outperform other players by modifying certain in-game elements through hardware or software to give the user unfair advantages. For example, aiming robots (aimbots), which automatically aim a gun at other players, are widely used by cheaters who play first-person shooters (FPS), in which players fire weapons such as firearms. Other cheaters use a so-called ‘wallhack’, which makes walls transparent and allows players to see or attack competitors hiding behind a wall.

Cheating in online gaming is analogous to using performance-enhancing drugs in professional sports or cheating on a test in school. Since no one wants to play in unfair games, cheaters make it difficult for fair players to continue playing games. A recent global survey of 5,911 gamers in six countries (China, Germany, Japan, South Korea, the United Kingdom, and the United States) found that 60% of gamers have had their multiplayer gaming experience negatively impacted by other players cheating on multiple occasions. (Irdeto Global Gaming Survey, 2018). In addition, 77% of respondents said they would be likely to quit games if they encountered cheaters. Cheating presents game companies with serious problems including damaged reputations and loss of revenue. Thus, many game companies attempt to detect cheaters by adopting anti-cheating solutions or soliciting reports from players who have encountered cheaters. They often punish cheaters by suspending or permanently banning their accounts. However, the problem of cheating is not simply solved because it is difficult to implement security measures that effectively block all types of cheating methods and new cheating tools that can circumvent current monitoring systems are continually being developed.

The issue of cheating is more problematic if cheating is contagious like an epidemic. Lofgren and Fefferman (2007) introduced the outbreak of a virtual plague that happened accidentally due to game design flaws in the *World of Warcraft*, which is one of the most popular massive multiplayer online role-playing games (MMORPGs). They suggested that online gaming might contribute to understanding the pattern of a spreading disease as a useful research tool. Beyond infectious diseases, their claim may be also true for behavioural contagion because players may be influenced by others' behaviour in virtual worlds. If the contagion of cheating exists in the game world, the first step to tackling the issue is to find the mechanism of contagion. Therefore, this study aims to uncover how cheating is contagious among players in an online gaming system by analysing gameplay logs of *PlayerUnknown's Battlegrounds* (hereafter *PUBG*), which is a popular online multiplayer FPS game.

In *PUBG*, up to 100 people fight each other to be the last-standing player by killing all other opponents. When playing in teamplay modes, the last team standing wins the round. At the beginning of a match, players start from different spots on the virtual battlefield. Players scavenge for weapons, which can be found all over the playing field, and kill other players they encounter to survive and win the match. Players can make a car or boat trip to move from one place to another. To prevent players from deliberately avoiding combat or hiding for a long time, the game constantly reduces the size of the battlefield over time. If players do not move to the safe area, they are harmed by a player-damaging barrier or bombs and eventually get killed. Thus, players ultimately end up in one place at the game's conclusion after moving constantly to avoid danger zones.

Studying FPS games is a meaningful endeavour with respect to social science research, due to the competitive nature of the games. FPS is a popular game genre in which players use weapons to kill other players' characters in the first-person point of view. It is notorious for attracting many cheaters because its nature is more intensely competitive compared to other genres. Players tend to have a strong competitive spirit and thus may be more tempted to use cheating tools when they play FPS games such as *PUBG*. Although PUBG Corporation, which operates the game, bans millions of cheaters and has taken legal action against cheat code providers (PUBG, 2019), the problem of cheating shows no sign of abating as noted by BattlEye, which provides anti-cheating measures for *PUBG*, on Twitter (BattlEye, 2018).

Aside from the fact that it is becoming difficult to ignore the problem of cheating in the online gaming industry, the rationale behind selecting cheating, rather than other types of antisocial behaviour, as a research subject is that it is easier to observe and measure by combining gameplay records with a list of banned cheaters provided by the game company. In addition, the game design of *PUBG* allows to quantify the amount of damage done to victims by cheating as explained in Section 5.2.

In recent years, previous research explored whether online game cheating spreads through friendship networks. Thus, more efforts are required to understand how cheating spreads among strangers to fill the gap in the current literature. In doing so, this study aims to test whether cheating spreads through contacts with competitors who use cheating tools. As the game arranges random matchmaking by default, players are grouped with strangers as competitors for each match. This feature of the game makes studying cheating in *PUBG* particularly useful in understanding how a player's decision to adopt unethical behaviour might be influenced by strangers. It may be possible that victims who have been killed by cheaters adopt cheating to retaliate against other players in the game world. As this hypothesis has not been addressed adequately enough with empirical data, this dissertation seeks to test whether players are more likely to engage in cheating if they have been harmed by cheaters in the game world.

To date, there has been relatively little social science research on cheating in online gaming. In contrast, many papers in the fields of engineering have attempted to detect cheating or implement security measures against cheating. Instead of focusing on technical protection measures for game systems, this paper aims to provide information that helps game companies regulate cheating by identifying high-risk players who are more likely to begin cheating. Understanding the underlying contagion mechanisms is important for regulating the spread of cheating because different approaches should be adopted to tackle different scenarios of contagion. For example, players who frequently team up with cheaters should be monitored if cheating spreads within friendship networks. If victims of cheating are more likely to become cheaters, they should be assigned to a high-risk group. Thus, the results of this research could help game companies identify potential cheaters.

The remaining part of the paper is organised as follows. Section 2 reviews the existing studies related to the topic and formulates a research hypothesis. Section 3 describes the data and data collection methodology used for this study. Section 4 is concerned with the analytic strategies which include a network motif analysis and data-driven estimation of missing information. Section 5 presents the findings of the research, focusing on the ‘generalised reciprocity’ mechanism that underlies the contagion of cheating. Section 6 discusses the implications of the key findings and includes limitations of this project. I also suggest future research ideas at the end of the paper.

2 Literature Review

2.1 The contagion of antisocial behaviour

Behavioural epidemics have been studied by many social scientists using the experimental method. Because little is known empirically about how antisocial behaviour such as cheating spreads in online games, this study relies on previous experimental research to establish the theoretical background of the topic.

Much of the literature has documented the peer effects on socially undesirable behaviour such as dishonesty, drug use, and aggression (Gino, Ayal, & Ariely, 2009; Gould & Kaplan, 2011; Jung, Busching, & Krahé, 2019). Previous studies have highlighted the role of peer influence in adolescents' antisocial or health-risk behaviours, such as smoking, alcohol use, and violence (Brechwald & Prinstein, 2011; Bond & Bushman, 2017). Although peer influences can be either positive or negative, Dimant (2019) demonstrated that antisocial behaviour spreads more strongly than prosocial behaviour among peers with a higher level of social proximity. In light of this evidence, it is now well established that antisocial behaviour spreads among friends.

However, peer influences cannot be attributed solely to the contagion of antisocial behaviour. While the topic of peer effects in the contagion of antisocial behaviour is not completely novel, it has been demonstrated that victimisation experiences or observations can also be important influencing factors on the spread of antisocial behaviour. Several studies have reported that victims of school bullying are more likely to become bullies as a result of their victimisation experiences (Han, Ma, Bang, & Song, 2019); similarly, people who witnessed or experienced domestic violence in their childhood later abuse children or become involved in general violence (Murrell, Christoff, & Henning, 2007). However, the mechanisms that underpin retaliation or social learning are not yet fully understood as these studies mainly explored the contagious aspect of violence that occurs in one’s social life over a long period.

Although it has been reported through field experiments that people who observe graffiti or littering are more likely to engage in stealing (Keizer, Lindenberg, & Steg, 2008), a systematic understanding of the potential contagion of antisocial behaviour through one-time contact between strangers is still lacking. In this regard, several studies have begun to explore the contagion of antisocial behaviour in the virtual world without face-to-face interactions. It has been recently observed that if people are exposed to so-called “troll” posts (or “trolling”), which intentionally cause conflict between people on an online news site, they have a greater chance of becoming a “troll” themselves (Cheng, Bernstein, Danescu-Niculescu-Mizil, & Leskovec, 2017). Kwon and Gruzd (2017) attempted to distinguish the effects of experience and observation of offensive comments on YouTube and found that a comment that contains profanity causes further profanities in others’ replies to the comment. However, the consequences for people who have been targeted or exposed to trolling and swearing remains unclear as it is difficult to quantify the extent of emotional damage that people receive resulting from each case.

Little research has been done on the causal mechanisms that underlie the contagion dynamics of antisocial behaviour. Tsvetkova and Macy (2015) distinguished the effects of victimisation and observation through an online experiment and suggested that antisocial behaviour can spread through two distinct mechanisms. The first, referred to as ‘generalised reciprocity’, indicates a case in which victims of antisocial behaviours seek to repay it to others. Although generalised reciprocity initially indicates a situation in which people who have benefited from a stranger’s prosocial behaviour seek to repay favours to others, it has been found that victims of antisocial behaviour also tend to retaliate against other people (i.e., A harms B; then B harms C), findings which may also be applied to computer-mediated deviance such as online game cheating. The other mechanism, known as ‘third-party influence’, describes a situation in which people learn antisocial behaviours by observing and imitating others. The authors found that people are less likely to harm others after observing low levels of antisocial behaviour. Focusing on the generalised reciprocity mechanism, this study examines the interactions between players in a hostile environment and explores how they shape the spreading pattern of cheating in the online game world of *PUBG*. It is hypothesised that a victim of cheating is more likely to become a cheater (i.e. cheater A kills non-cheater B; then B adopts cheating and kills C).

2.2 The contagion of cheating in the online gaming context

Although cheating has been the subject of many studies in the fields of sports and education studies, research on cheating in online gaming is scarce. Some researchers interviewed or surveyed cheaters, asking them about the motivating reasons why they cheat. Wu and Chen (2013) found that players are more likely to cheat if they expect to benefit from cheating or do not take cheating seriously. Chen and Ong (2018) investigated the psychological processes that cheaters adopt to rationalise cheating. Both studies discovered that exposure to friends who cheat is also closely related to cheating adoption of non-cheaters. While these qualitative studies have revealed various innate factors of cheating adoption, little quantitative analysis of online game cheating has been conducted. Thus, this study employs a quantitative approach to analyse gameplay logs provided by the game company. With a large-scale dataset, it is possible to quantify interactions and relationships between players and explore population processes, such as spread dynamics. Part of the aim of this project is to prove that large-scale game data could show promise of a better understanding of human behaviour in the online universe.

Only recently have researchers empirically investigated the contagion of online game cheating.

Blackburn et al. (2014) first identified online game cheating as a contagion issue and found that, for a non-cheating player, the number of cheating friends could be a predictor of future cheating adoption on the online social network of Steam, a digital game distribution platform where players can purchase games and make friends. Zuo, Gandy, Skvoretz, and Iamnitchi (2016) also reported that non-cheaters are more likely to adopt cheating if they are exposed to unpunished cheating neighbours in the Steam Community or to cheating teammates in the game. These studies support the hypothesis that players who have many cheating friends may start cheating as well. However, they made no attempt to distinguish the effects of social influence and homophily. In this regard, Woo, Kang, Kim, and Park (2018) attempted to distinguish the correlated behaviour and social influence by using a shuffle test as suggested by Anagnostopoulos, Kumar, and Mahdian (2008). They found that cheating is socially contagious among friends in an MMORPG. However, as mentioned in Section 2.1, unethical behaviour can spread among strangers through negative interactions. Since players typically cooperate with each other and maintain close relationships between friends in MMORPGs, studying MMORPGs is not appropriate to examine hostile interactions between players. As players basically kill each other in FPS games, this study examines whether non-cheating players begin cheating if they are harmed by cheating opponents in the FPS game.

Since previous works focused primarily on the contagion within friendship networks, much remains unknown about how cheating spreads through contacts with strangers in the game world. Only the study by Zuo et al. (2016) compared the fraction of players who began cheating after playing with cheating teammates versus those who played with cheating competitors. According to their findings, cheating teammates are more influential than cheating competitors as they are more likely to encourage non-cheaters to adopt cheating. However, instead of simply comparing the fraction of players who newly have begun cheating in the data, this study employs a network analysis approach and tests statistical significance to examine the effects of cheating opponents. I also distinguish the effects of mere exposure and victimisation by considering the amount of damage done to victims of cheating.

3 Data

3.1 Data sources

The data were gathered from two online sources between March 1 and March 31, 2019. First, this study obtained gameplay logs from the Korean server through application programming interfaces (APIs) for a period of one month. Steam and Kakao Games are two different platforms that currently distribute the game in South Korea (hereafter Korea). Among these platforms, this study employs the data retrieved from the server of Kakao Games because the game publisher uploads a list of banned cheaters on its website every day, thus facilitating the identification of cheaters.

It is challenging to define online game cheating because cheating methods vary by game genre and new ways of cheating are constantly being developed. Moreover, it is impossible to identify cheating tools adopted by banned cheaters based solely on gameplay logs. Since it is not the task of this paper to define online game cheating, this paper only relies on a list of cheaters shared by the game company when identifying cheaters among players. In this regard, cheating refers to any usage of third-party software that violates the policies of the game company throughout this study.

Another advantage of choosing the server that Kakao Games operates in Korea is that it has a small population compared to the servers maintained by Steam. The rationale for choosing the server with a smaller population is that this study ideally aims to construct a network that has similar properties to the real network in order to trace the effects of influence accurately. It is difficult to obtain a network sample that preserves the underlying properties of the real network if the size of a population is too large. Although different methods of network sampling such as node sampling, link sampling, and snowball sampling are widely used, it is important to choose an appropriate method for the network metric of interest because there is no single unbiased sampling method for all network properties (Lee, Kim, & Jeong, 2006). However, these methods show less variation in outcomes as the sampling fraction, which is the ratio of the number of sampled nodes to the total number of nodes in the original network, increases. In order to increase the sampling fraction, a relatively small network is selected in an effort to capture as many nodes and links as possible and thus obtain a realistic network.

Due to restricted access to the data, this study employs a breadth-first search (BFS) sampling. Although the structure of the entire network is hidden, it is possible to explore the neighbours of a node by using the APIs. To elaborate on the sampling process, a set of players was first retrieved as seed players who participated in a set of random matches returned by the APIs. At the first iteration, players who took part in the same matches with the seed players and a set of matches played by them were discovered. With newly retrieved matches, the process above was repeated until all new nodes (players) who played the same matches together were discovered.

The figure below illustrates an example of the sampling process. In Fig. 1, pairs of players are connected if they participated in the same match. The process starts at the zero state with a seed player. Players labelled as one are the nodes collected at the first phase of the process and players labelled as two are the nodes collected at the second phase, and so forth. Since the out-of-date logs expire after a retention period of 14 days, it was necessary to revisit logs of all players repeatedly in order to obtain longitudinal data during the observation period.

A total of 1,291,441 matches played during March were collected initially. The data for each match has a link to its telemetry file, which contains a list of events that happened during the match and detailed logs. However, in the data collection process, it was impossible to decode some telemetry files due to a few technical issues with the APIs. Among 1,291,441 matches, 91% had their telemetry files. This study omitted matches without telemetry files, and this results in the final dataset comprising a total of 1,179,537 matches with telemetry files.

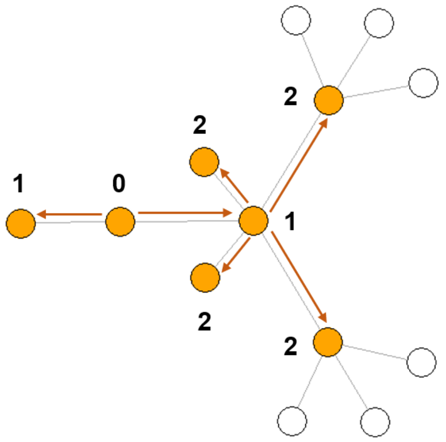

Fig. 1. Example of a network sampling (Sequence of nodes collected during two iterations)

3.2. Temporal network construction

This study uses networks to represent interactions between killers in the game and their victims. A time-stamped directed edge from player A to player B means that A killed B at a certain time during the match. Data of each player basically contain a nickname and a player ID. If a player has been banned due to cheating, the corresponding player's profile additionally contains a ban date. Each edge has a match ID, an attack ID which indicates unique contact between two players, and a timestamp which records when the attack happened.

Fig. 2 illustrates an example of a killing network constructed from a squad match. Red nodes and blue nodes indicate cheaters and non-cheaters respectively. All self-loops, which point to themselves, are removed from the network because they are not interactions between killers and their victims. Although not mentioned by the PUBG API document, it is likely that self-loops are players who killed themselves or were killed in accidents, not in combat. All contacts between players from 1,179,537 matches are aggregated into a large directed multigraph where pairs of two players can have multiple edges between them. The observed killing network contains a total of 100,927,444 edges between players.

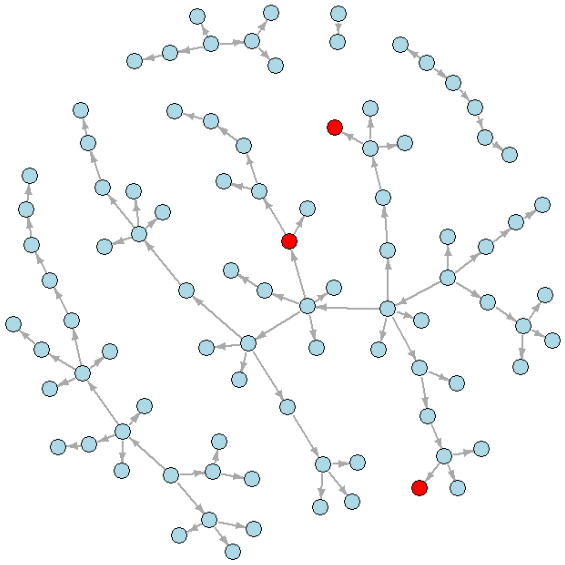

Fig. 2. Network visualisation of a killing network constructed from a squad match

4 Methods

4.1 Motif analysis

Network motifs are subgraph patterns with a fixed topology that occur more frequently in the real network than in random networks (Milo et al., 2002; Holme & Saramäki, 2012) and have been used to understand complex networks in a variety of research areas. Unlike static snapshot networks, temporal networks with time-stamped edges include the timing information of events. Thus, temporal motifs are topologically equivalent patterns that also include the same order of events and are defined to occur within a specified period of time. Several studies have revealed the dynamics of social systems by using a temporal motif analysis within the field of social science (Tsvetkova, García-Gavilanes, & Yasseri, 2016; Kovanen et al., 2011).

Many researchers have attempted to identify motifs that emerge from data or find the most frequent motifs by implementing fast and efficient motif detection algorithms (Srivastava, 2010; Kovanen et al., 2011; Paranjape, Benson, & Leskovec, 2017). In this study, network motifs in a temporal setting are used to represent underlying dynamical mechanisms. Thus, the network motif is the case where a non-cheater begins cheating after being harmed by a cheater. This study tests whether the frequency of the given motif in the observed network is significantly different compared to conditionally randomised networks.

The motif analysis approach was chosen because it is possible to confirm the presence of the function or structure of the complex system and provide statistical significance of contagion by testing whether the overrepresentation of a certain motif occurs due to mere chance or not (Milo et al., 2002). Using the timing information of temporal motifs, it is also possible to set a time window of observation when cheaters might have an influence on non-cheaters. This study assumes that the contagion effect operates for the transitions from non-cheaters to cheaters that occurred within a period of seven days. For example, it is less likely, due to the contagion effect, that a non-cheater adopts cheating one month after contact with a cheater.

To construct the null model based on randomisation of the empirical data, this study employs a node-label permutation approach, which preserves the structure but permutes the node labels of the network (Croft, Madden, Franks, & James, 2011). The nodes of each match are relabelled on the observed network to make the metric of interest random. This study performs the process 100 times and counts the number of instances of the given motif on each randomised network. As a final step, the z-score is estimated to assess statistical significance in the following way:

$$Z = (N_D - \mu(N_R))/\sigma(N_R). \tag{1}$$

In Equation 1, N_D is the count of motifs in the data. The mean and standard deviation of the count of motifs in the randomised networks are denoted as $\mu(N_R)$ and $\sigma(N_R)$ respectively. The z-score is used as a measure of position in the distribution of test statistics to test whether the observed metric is extreme. In other words, this shows whether there is a significant difference in the count of motifs between the empirical network and node permuting null model exists. The null hypothesis is that the count of motifs on the empirical network is no different from random. All analyses including the construction of randomised networks were carried out using Spark and Python.

4.2 The contagion of experienced cheating among strangers

This study counts pairs of players who show the given motif as a test statistic. To test whether cheating spreads through contacts with strangers, this study suggests a 2-node temporal motif as a measure. To examine whether cheating spreads through a victimisation experience among strangers, this study compares the observed network and reference networks based on randomisation of the node labels. Fig. 3 below illustrates the network motif that demonstrates the victimisation-based mechanism. In this study, the temporal motif is a representation of the case where a non-cheater adopts cheating after being killed by a cheater within a specific time limit, which is seven days. The time difference between the contact with a cheater and the transition from non-cheater to cheater should be no longer than seven days. If non-cheaters transition to cheaters within the period of seven days, they are considered to be influenced by cheaters.

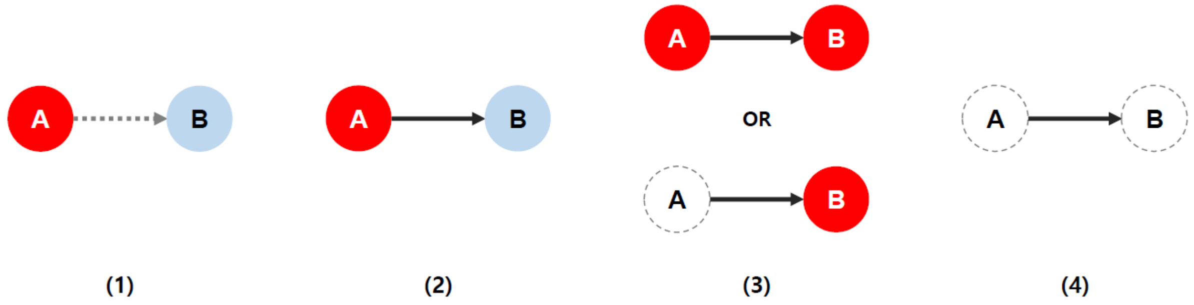

Fig. 3. Example of a 2-node temporal motif as a representation of the victimisation-based mechanism. If a non-cheater adopts cheating, its colour becomes red. Dotted circles indicate that cheaters are banned due to cheating. The edge between two players becomes active at the stamped time. The dashed edge is not currently active but is meant to be formed later, whereas black edges are already active. The order of events is as follows: (1) Player A starts cheating before killing player B; (2) A kills B; (3) B adopts cheating after being killed by A; (4) Both of them eventually get banned. Note that A could be either undetected or banned at the time when B begins cheating.

To confirm that the count of motifs has not arisen by chance in the observed network under a null hypothesis, comparisons are made between the observed network and randomised networks. It is important to design null models that have realistic features except for the metric of interest, controlling for other network properties. As explained in the following section, a large gap exists between cheaters and non-cheaters in terms of performance. Cheaters tend to kill more opponents than non-cheaters and thus show a higher outdegree. In this case, simply permuting all node labels without considering node type does not generate an appropriate null model as it could lower the outdegrees of cheaters and eventually make cheaters kill fewer players than they actually do in the real network. The above approach reduces the outdegrees of cheaters and consequently affects the count of motifs. Thus, this study preserves the outdegree distribution of cheaters and that of non-cheaters as the empirical network by permuting the node labels within each match. For each match, node labels are re-allocated within cheaters and within non-cheaters separately. Another rationale for choosing node permutations within matches is that the observed network involves temporal constraints. Players tend to play the game only for a certain time period as shown in Section 5.1 and cheaters are removed from the game after they are banned. Thus, simply shuffling edges is also not appropriate in this case because it fails to control for temporal features of the killing network.

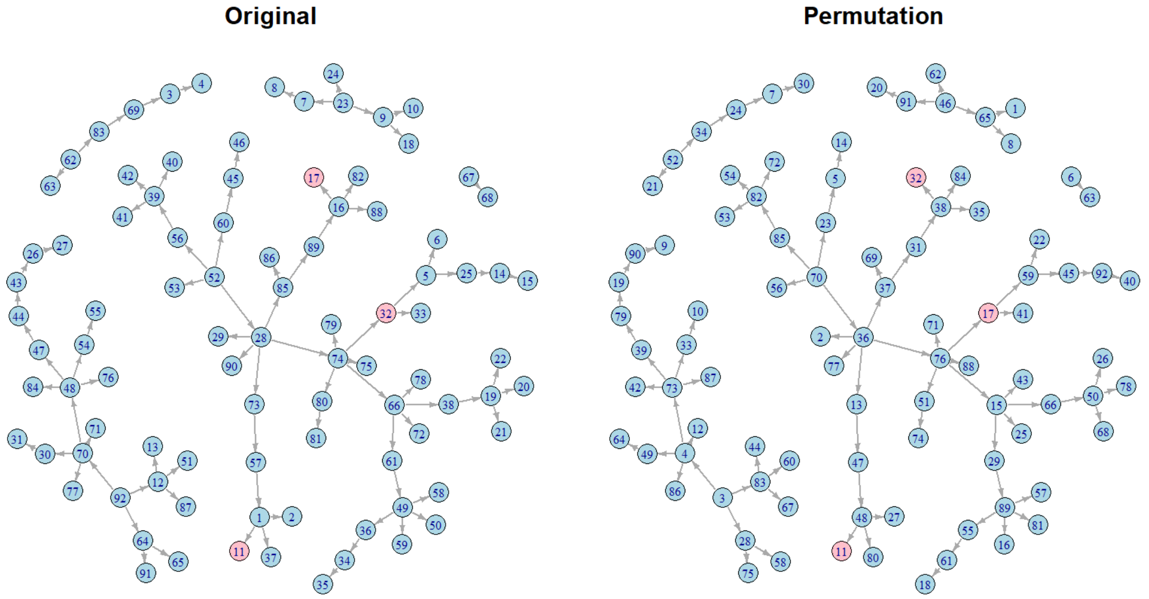

Fig. 4. Example of a node permutation on the killing network constructed from a squad match

The process of a node permutation is illustrated in Fig. 4. Cheaters are in the same positions on the permuted network as they used to be in the original network, but the node labels of cheaters can be swapped with each other. It should be noted that among 130,540 matches in which at least one cheater took part, 112,799 matches had only one cheater (see Fig. 3 in Appendix). Thus, in most cases, cheaters remain at the same position in which they had been on the original network. Swapping the node labels of non-cheaters happens only between non-cheaters. The process of permuting the node labels within matches preserves the network structure consistency including time constraints while allowing the victimisation experiences of players to change. After constructing reference networks with permuted node labels, the next step is to count how many times the temporal motif occurs on each network with a time window of seven days. Finally, the observed test statistic is compared to the values of randomised networks in order to test whether it is significantly different from random. In other words, this study tests whether the given motif pattern appears more frequently in the empirical network than in randomised networks.

4.4 Estimation of the duration of time for cheating adoption

One of the most common issues that studies face in assessing the contagion of cheating in online gaming is the lack of information about the duration of time for cheating adoption. Because the game company only provides the date when a ban was applied to each cheater, this study attempts to estimate the time when cheaters started cheating. Prior to analysing the data, this process is required to identify the order of actions and who influenced whom. Although the game company tries to detect cheaters as soon as possible, some cheaters are reported or detected quicker than others, whereas others may go undetected for longer time periods. In some cases, the game company needs additional time for a thorough inspection before banning cheaters. For these reasons, it is likely that cheaters began cheating at different times even if they were banned on the same day.

To compensate for this missing information, this study uses gameplay logs to estimate when a cheater started to cheat. Past research has not taken full advantage of gameplay logs to tackle this issue. For example, Zuo et al. (2016) assumed that cheaters had always cheated before they were banned. A major disadvantage of the above approach is that it overestimates the impact of cheaters. To counter the potential problem of overestimation, this study employs a data-driven approach by using gameplay logs and features that differentiate cheaters and non-cheaters in an attempt to estimate the timing of cheating adoption more accurately.

Up to now, different methods that use gameplay logs have been proposed to detect cheaters. Many studies have found that the features which are related to kill scores detect cheaters effectively in FPS games. This paper employs a player behaviour analysis approach, which is based on the idea that cheaters behave differently from non-cheaters (Alayed, Frangoudes, & Neuman, 2013). To distinguish cheaters from non-cheaters, the features presented in Table 1, which were suggested in a study by Park, Han, and Kim (2015), were used. The study uses the features in the table to estimate the time between when cheaters started cheating and when they were banned. It should be noted that there may be some cheaters who performed poorly even if they use cheating tools. In these cases, it is difficult to detect the timing of cheating adoption because the proposed player behaviour analysis assumes that cheaters perform better after they start cheating. However, this study still argues that the estimation should be based on kill logs because it is likely that non-cheaters are tempted to cheat when they observe that cheating tools are effective; whereas if non-cheaters observe that cheaters perform poorly, they will not be motivated to cheat. Thus, this study assumes that the more cheaters are successful in killing other players, the more influential they are. In other words, performance is regarded as a proxy for the level of influence a cheater can exert on non-cheaters.

Table 1: Description of features

Feature	Description
Average kill ratio per day	Average ratio that a player kills opponents in a day
Average time difference between consecutive kills per day	Average of time difference between two consecutive kills in a day
Difference between the recent kill ratio and average kill ratio	Difference between the most recent kill ratio of a player in a day and average kill ratio

First, the *average kill ratio per day* is calculated by dividing the number of kills by the sum of the total number of kills and that of deaths in a day. The *average time difference between consecutive kills* is measured if a player killed at least two other players during the match. This measure serves as an important feature for cheating detection. In general, cheaters tend to show abnormal and noticeable results with the help of cheating tools. In *PUBG*, all players can see who killed whom in real-time through the “kill feeds” that appear on the upper right of the screen during the match. Some players have recognised the presence of cheaters by viewing the kill feeds that display killers and their victims on a real-time basis, as these logs may reveal bot-like behavioural patterns of cheaters. For example, the player who kills many other players consecutively in very short time intervals is more likely to be suspected as a cheater. Finally, the *difference between the recent kill ratio and average kill ratio* is measured by subtracting the most recent kill ratio per day to the overall average kill ratio. It has been found that this feature effectively differentiates between cheaters and highly skilful non-cheaters (Park, Han, & Kim, 2015).

To confirm the idea that cheaters and non-cheaters are fundamentally different in terms of performance, this study first calculates the average kill ratio to compare the two groups. In this section, this study tentatively supposes that cheaters who were banned between March 1 and March 3 always cheated for use as a baseline for the estimation. The number of cheaters who were identified during this period was 651 and 854,877 non-cheaters played at least one match during the same period.

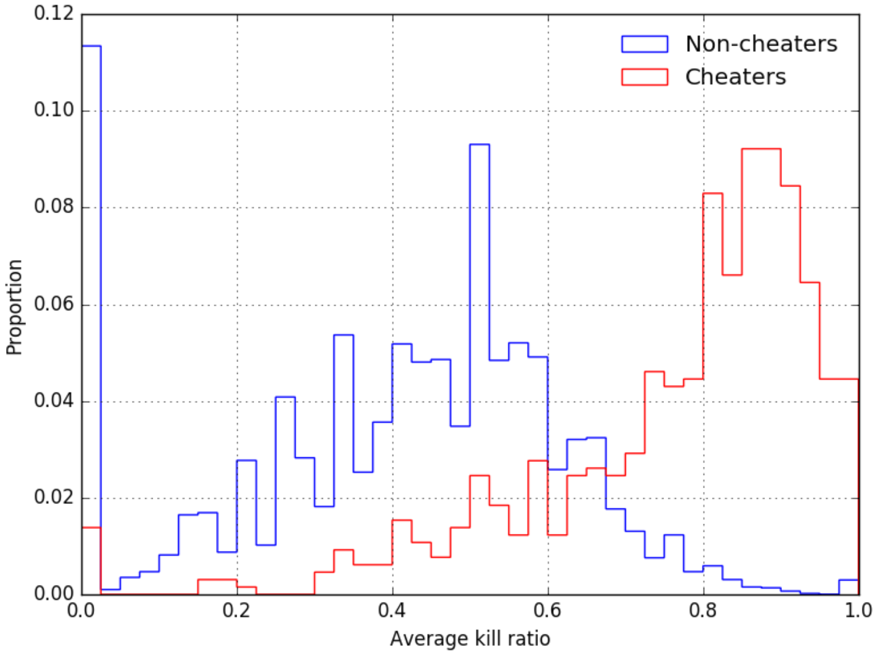

Fig. 5. Proportion of the average kill ratio between March 1 and March 3

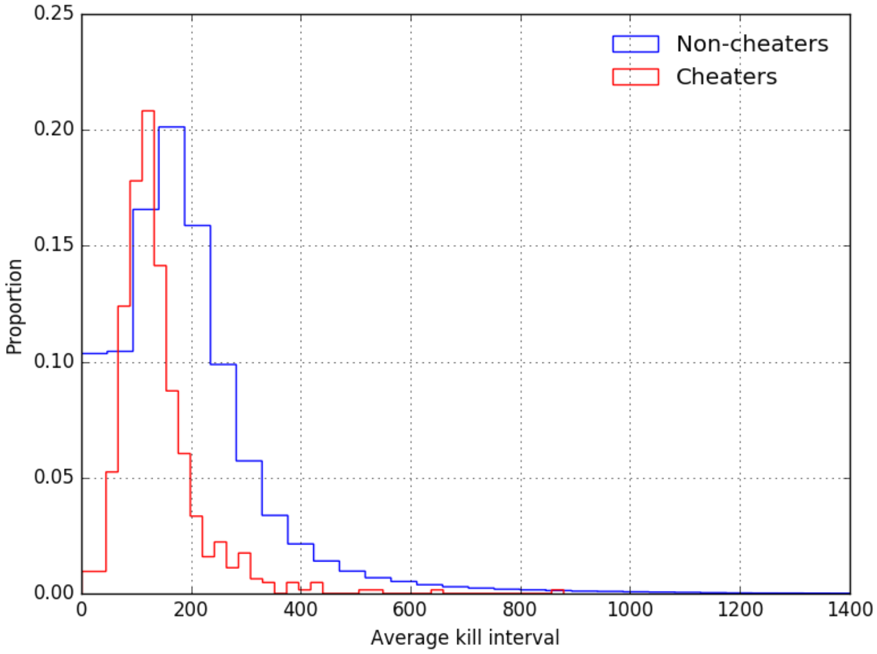

Fig. 6. Proportion of the average time difference between consecutive kills between March 1 and March 3 (with outliers excluded)

According to Fig. 6, cheaters perform better than other players as they show a higher average kill ratio. The mean values of cheaters and non-cheaters are 0.77 and 0.41 respectively. The median value of cheaters is 0.82 and that of non-cheaters is 0.44. Using two independent sample *t*-tests, comparisons between the two groups were made. The difference in the average kill ratio between the two groups is significant, $t(855{,}526) = 48.41, p < 0.01$.

The overall average time difference between kills was shorter for cheaters than non-cheaters as shown in Fig. 7. The mean values of cheaters and non-cheaters are 139.17 and 192.27 respectively. The median value of cheaters is 122.64 and that of non-cheaters is 170.33. There is also a significant difference in the average time difference between consecutive kills, $t(629{,}553) = -17.44, p < 0.01$.

After tracking changes over time, Fig. 1 in Appendix presents the values of all cheaters on the change in the average kill ratio. It can be seen from the figure that the average kill ratio change tends to be above zero for the majority of cheaters, thus indicating that cheaters gain advantages after adopting cheating methods and kill more opponents than they had done previously.

For each cheater in the dataset, the start date of cheating was estimated by using these three features. On the basis of the results, this study assumes that cheaters started cheating beginning on the date when they met at least two of the following conditions: 1) average kill ratio greater than or equal to 0.8; 2) average time difference between consecutive kills was equal to or shorter than 140; and 3) average change in kill ratio was greater than or equal to 0.1.

Among 6,161 identified cheaters, complete performance information was available for 3,646. According to the estimation results shown in Fig. 7, the period of cheating varied. They did cheat for five days on average and the modal value of the period was two days. For 2,515 cheaters, who had at least one missing value on performance or did not meet the conditions above, the modal value of two days was applied as the period of cheating. In the end, around

62% of all cheaters (3,850 players) were estimated to have cheated for two days. After estimating the timing information, the next step is to identify who cheated first and then who influenced whom.

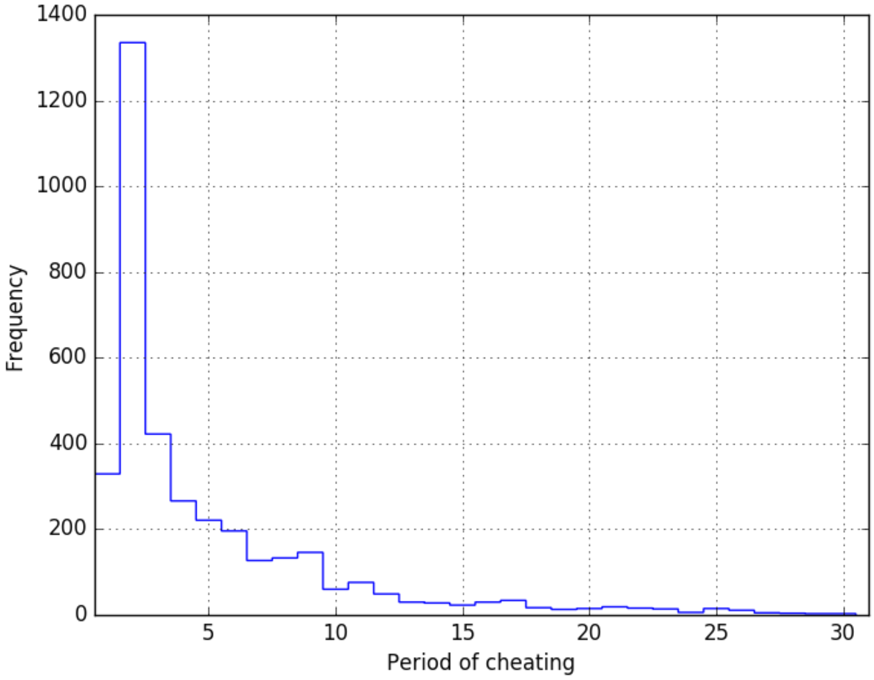

Fig. 7. Histogram of the period of cheating of 3,646 cheaters who had full information on performance

5 Results

5.1 General statistics

The dataset contains 1,179,537 matches played during the observation period, including 1,985,771 unique players and 6,161 cheaters who comprised 3.1% of all players. The number of matches played per day peaked every weekend (see Fig. 2 in Appendix). By counting the number of days when a player had at least one match record, it has been found that most players played less than five days as shown in Fig. 8. The median of days players played the game was three days during the observation period. It can be seen from the figure that only a small portion of players played the game regularly every day, whereas most players accessed the game for a limited period of time.

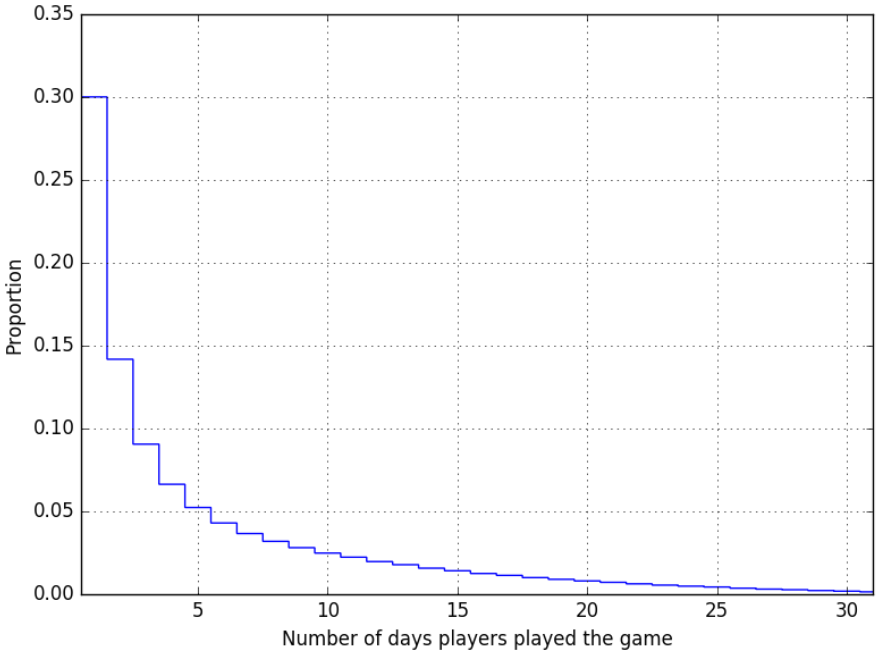

Fig. 8. Proportion of the number of days *PUBG* players accessed the game during the observation period

The game provides modes in which players cooperate with each other to fight against other teams. In *PUBG*, there are three different game modes: solo, duo, and squad. When playing a solo match, players are on their own without teammates. However, there are two gameplay modes in which players are grouped into teams. In teamplay modes, players can invite their friends to form a team and players who do not have friends to invite are matched with random teammates. Up to two players can form a team when they play a duo match and up to four players can play a squad match as a team. As presented in Table 2, players tend to play more duo or squad matches than solo matches showing a preference for team play than solo play.

Table 2. Number of matches by game mode

Game mode	Number of matches
Solo	132,572 (11.2%)
Duo	370,674 (31.4%)
Squad	676,291 (57.3%)
Total	1,179,537 (100.0%)

Table 3. Proportion of winning or getting into the top 30 percent of teams with a different number of cheating teammates.

Number of cheating teammates	Number of teams	Proportion of wins	Proportion of top 30 percent
0	33,677,313	0.03	0.64
1	121,872	0.13	0.87
2	5,213	0.24	0.90
3	299	0.39	1.00
4	22	0.68	1.00

The abnormal performance of cheaters can be seen even in terms of team performance. Table 3 shows how the results of matches vary dramatically by the number of cheating teammates. Only teams of at least two players are considered in Table 3. All individuals who played teamplay matches alone are excluded from the table. It can be seen that the chance of winning or getting into the top 30 percent of all teams becomes staggeringly high with an additional cheating teammate. Thus, cheating determines whether the team gets to the top or not.

5.2 The contagion of experienced cheating among strangers

Turning now to the contagion of cheating among strangers, the present research explores the effects of cheating competitors on the spread of cheating and tests whether victims of cheaters are more likely to become cheaters. During the observation period, 311,311 players were killed by cheaters at least once. The distribution of *z*-scores is plotted to test whether the count of motifs with a time window of seven days on the empirical network is significantly extreme compared to randomised networks. If cheating does not spread through contacts with cheating opponents, the count of motifs on the empirical network should be similar to that on the null model. When the amount of harm done to victims is not considered, Fig. 9 shows that no significant differences were found between the observed network and randomised networks.

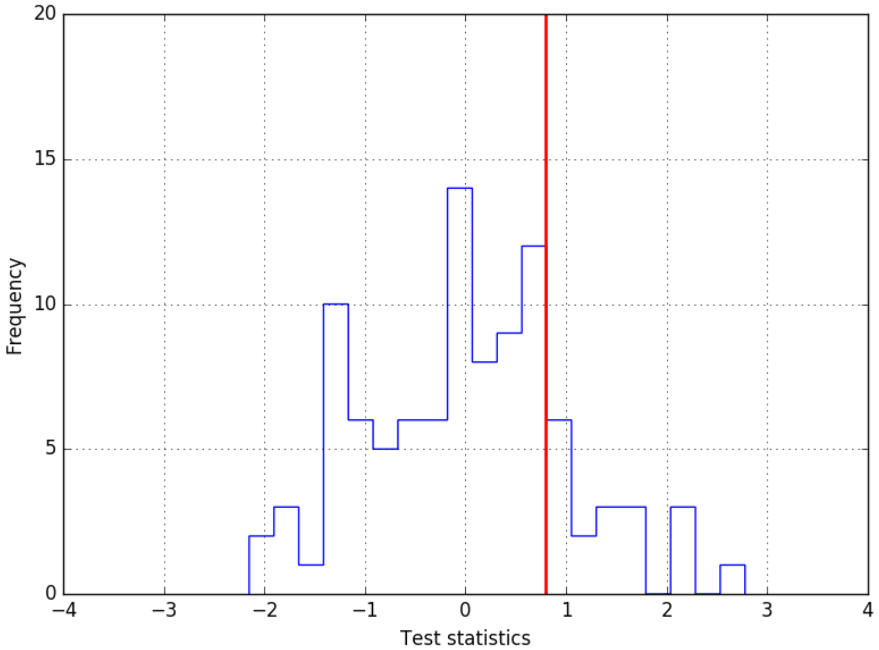

Fig. 9. Distribution of test results without considering the severity of damage done to victims

Because up to 100 players participate in a match, it is difficult to rank first in *PUBG*. Thus, players who got into the top 10 or 30 percent of all participants are also considered to have performed well in the match. As a typical battle royale game, players are given only one life per match in *PUBG*. The game does not allow players to respawn and any players who die lose the match. This system maximises the competitive nature of *PUBG*. Thus, a strong possibility exists that players who are killed by cheaters after surviving long enough to be ranked in the top 30 percent feel more frustrated than those who are killed early. In this regard, this study excluded 32,561 matches in event modes, which allow players to revive multiple times during the match, out of the total matches. A possible explanation of why the results are not significant when the amount of damage done to victims is not considered is as follows: Players who are killed early may be poor performers who are often eliminated early regardless of whether the competitor is cheating or not. Even if skilful non-cheating players are killed by cheaters in an early phase of the match, they might consider it as negligible damage and just begin a new match.

Thus, this study tests whether the level of harm caused by cheating affects the likelihood of a non-cheater's adoption of cheating. In this case, the network motif is the case in which a non-cheater begins cheating after being killed by cheating after they have entered the top 30 percent during the match. It can be seen from the results in Fig. 10 that non-cheaters are influenced by cheating competitors when they have been severely harmed. The observed network shows a significantly higher frequency of the motif compared to randomised networks. As competition becomes intense and only skilful players tend to survive over time, even minor cheating can change the result of a match. Thus, it is likely that players who have performed well and survived longer are more susceptible to cheating and retaliation. Fig. 11 shows the difference in time between the starting date of cheating and the time when a player was killed by a cheater for the first time after reaching the top 30 percent. However, Fig. 11 should be interpreted with a caveat: There is no value for hours, minutes, or seconds for the starting date of cheating because it is defined by day.

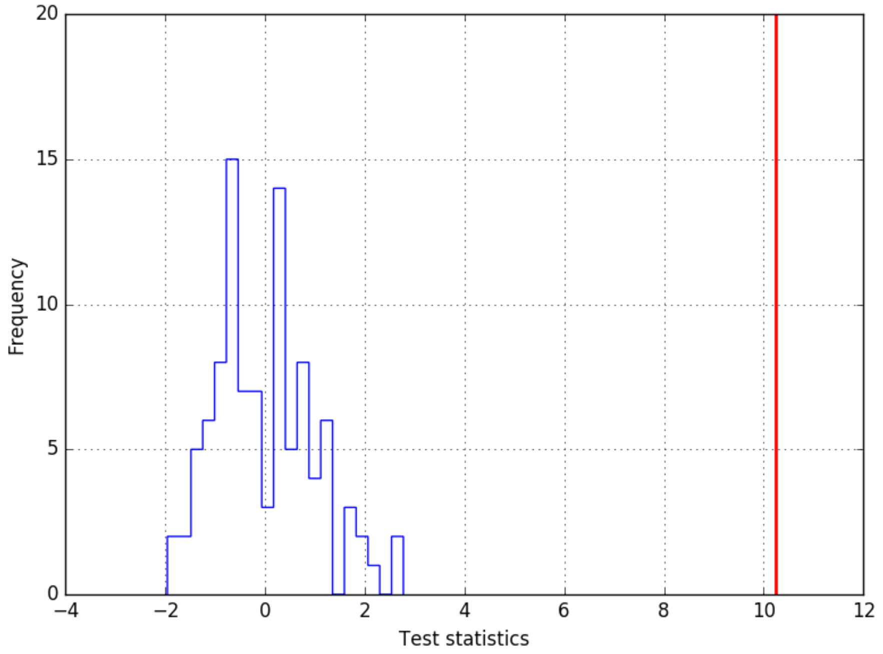

Fig. 10. Distribution of test results considering the severity of damage done to victims

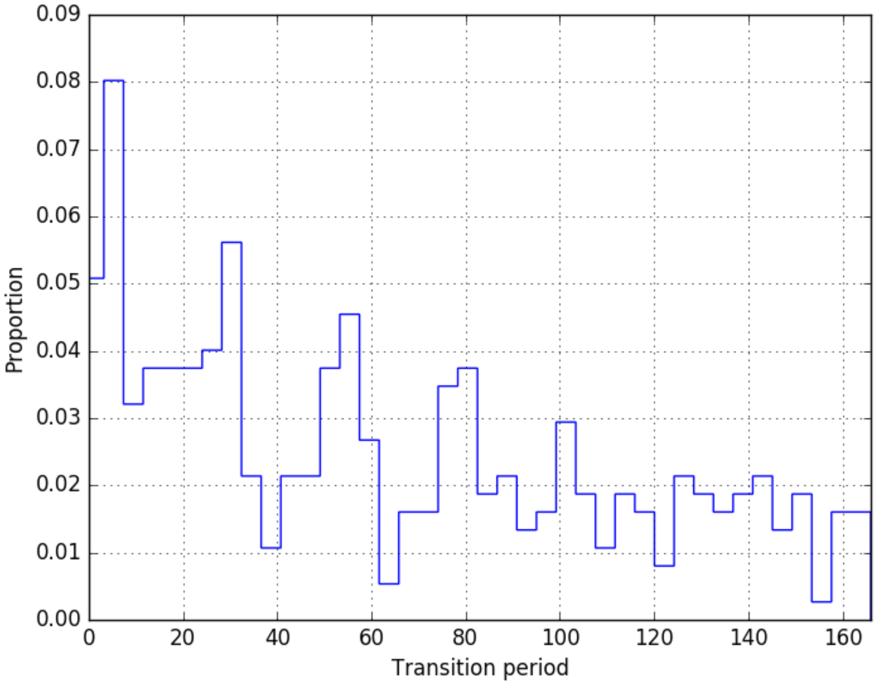

Fig. 11. Distribution of time difference between the transition from non-cheater to cheater in

hours

In summary, this finding reveals that the victimisation-based mechanism is embedded in the empirical data and shapes the spreading pattern of cheating among strangers. The result accords with the experimental evidence on the generalised reciprocity mechanism found in the previous study. Additionally, it is a new finding that the level of damage done to a victim is a significant factor in the generalised reciprocity mechanism. It has been suggested that the likelihood of adopting unethical behaviour decreases in the presence of penalties (Zuo et al., 2016; Gould & Kaplan, 2011). Kakao Games has a strict policy against cheating in *PUBG*: Cheaters who have been banned are no longer able to purchase another copy of the game. However, the evidence indicates that the desire for retaliation is strong enough to spread like a contagious disease even when the punishment is harsh.

Some cheaters rationalise their cheating behaviours by insisting that they cheat just for fun and that no one is harmed in the real world. Contrary to what cheaters believe, the outcome of their cheating arouses anger among fair players who devote time and effort into the game. The results above imply that players are willing to behave antisocially when they have been harmed and treated unfairly even if the antisocial action was made in the virtual world.

6 Conclusion

The main goal of the present study was to empirically test whether cheating spreads through contact with cheating opponents. This study analysed a large network of interactions between killers and their victims and tested the generalised reciprocity mechanism by using a network motif analysis. The results suggest that non-cheaters who have been severely harmed by cheaters are more likely to transition from fair players to cheaters. However, no significant difference existed between the observed network and randomised networks when the extent to the damage done to victims was not considered. Taken together, these findings suggest that the extent of the damage inflicted on victims plays an important role in the desire for retaliation. Victims do not adopt cheating simply because they have been killed by cheaters. Rather, they are more likely to begin cheating and retaliate against other players if they have received a significant amount of damage. Although results from qualitative studies have identified psychological processes that stimulate players to cheat, an implication of this study is that cheating is a social phenomenon and how players interact with cheaters also affects the likelihood of a non-cheater's cheating adoption.

Though the findings of this study are new and compelling, this study rests on some limitations. First, we do not know whether victims of cheaters were fully aware that they were harmed by cheaters because it is sometimes difficult to distinguish between cheaters and very skilful players. However, cheaters in FPS games show abnormal performance that often stands out. Moreover, *PUBG* has an in-game system which allows players to watch replay videos or watch the rest of the game from the perspective of their killers. The game company is currently using this system to encourage players to report cheaters and players actively use the system when they suspect that they may have been killed by cheating. Thus, with the help of the system above, this study assumes that victims of cheating recognise that they were killed by cheaters.

Secondly, because the study sample consists only of players who play on the Korean server, the results of this paper cannot be generalised to a broader population of players who play the game in different regions. Players in different regions may have different attitudes or susceptibilities to cheating and these characteristics could affect the contagion process of cheating. Another issue that was not addressed in this study is that players can create multiple accounts. However, Kakao Games, the publisher of *PUBG*, has strict rules as it requires players to verify their identities when they register for the game and allows only one account per player. Once cheaters are banned, they cannot purchase a new copy of the game for their banned accounts. Despite these strict rules, it is still possible for cheaters to create multiple accounts if they use identities of other people such as family members or friends. For this study, I assumed that each player had only one account. Since it is not possible to find out whether cheaters have multiple accounts, this paper should be interpreted with this caveat in mind.

In further investigation, the ‘third-party influence’ mechanism could be examined to test whether observed cheating spreads in online gaming. The contagion of observed cheating remains to be elucidated in the case of online gaming. A further study might conduct a more robust analysis with the use of spatial data. It is possible to retrieve location information including the coordinates of each player on a map of a battlefield on a real-time basis from a telemetry file for each match. Although some players may notice the presence of cheaters with the help of the kill feeds as mentioned in Section 4.3, many players may not notice until they actually encounter cheaters because players are scattered on a massive virtual battlefield. In order to identify exactly who observes whom, further data collection is required to measure the distance between players.

Although this paper did not study friendship networks in *PUBG*, a study that examines the contagion of cheating among friends could be undertaken once the definition of friendship is established. Moreover, the specific game system of *PUBG* raises a question regarding cheaters who are not purely being selfish. A further longitudinal study could observe whether non-cheaters ostracise or get along with cheaters. Blackburn et al. (2014) found that cheaters lose friends after they are labelled as cheaters in the Steam Community where players create their profiles and make friends. The paper concluded that non-cheaters tend to maintain the quality of their lists of friends by unfollowing cheaters. In the case of Steam Community, players do not necessarily need to own or play the same games to become friends. Thus, non-cheaters unfollow cheaters because befriending cheaters is not beneficial; rather, it only harms their own reputations. In contrast, non-cheaters can share benefits if they win a match with cheating teammates in *PUBG* as shown in Table 3. Even if a player fails to survive, it is still possible to win unless all teammates die because the last-standing team wins the round in teamplay matches in *PUBG*. Thus, non-cheating players have a better chance of winning the round by playing with cheating friends. Unlike in the Steam Community, it may be possible that cheaters are not socially penalised by non-cheaters in *PUBG* because non-cheaters might show a generous attitude towards their cheating teammates who are loyal to the team. Therefore, a focus on friendship networks could produce interesting findings that account more for cheating that favour other in-group members. A precise definition of friendship would give an opportunity to test a model of peer effects or explore a rationalisation of deviance that benefits in-group members.

The findings of this research shed new light on antisocial behaviour in computer-mediated contexts in a broad sense. Through the discovery of the contagion process of online game cheating, this paper also shows how social scientists can contribute to a better and fairer online gaming system for players. This project helps game companies monitor players who are at a much greater risk of cheating. The findings suggest that a reasonable approach is to monitor victims of cheating who have been severely harmed as a high-risk group. This information may also be applicable to many real-world cases of antisocial behaviour that feature aspects of competition, not only for online games.

Another contribution of this study is to prove that online gaming systems represent a useful research tool for social scientists to investigate how people behave and how their behaviours spread, especially in online environments. In other words, the online game world can be considered as the proxy of a larger social system. Large-scale digital records of player interactions and activities facilitate the understanding of population dynamics in ways that were impossible previously. It is hoped that this project will contribute to a better understanding of the mechanism by which antisocial actions permeate social systems.

References

Alayed, H., Frangoudes, F., & Neuman, C. (2013, August). Behavioral-based cheating detection in online first person shooters using machine learning techniques. In *2013 IEEE Conference on Computational Inteligence in Games* (pp. 1-8). IEEE.

Anagnostopoulos, A., Kumar, R., & Mahdian, M. (2008, August). Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 7-15). ACM.

BattlEye. (2018, February 4). We have banned over 1,044,000 PUBG cheaters in January alone, unfortunately things continue to escalate. [Tweet]. Retrieved from: https://twitter.com/TheBattlEye/status/960278229566226437.

Blackburn, J., Kourtellis, N., Skvoretz, J., Ripeanu, M., & Iamnitchi, A. (2014). Cheating in online games: A social network perspective. *ACM Transactions on Internet Technology*, 13(3), 9.

Bond, R. M., & Bushman, B. J. (2017). The contagious spread of violence among US adolescents through social networks. *American journal of public health*, 107(2), 288-294.

Brechwald, W. A., & Prinstein, M. J. (2011). Beyond homophily: A decade of advances in understanding peer influence processes. *Journal of Research on Adolescence*, 21(1), 166-179.

Chen, V. H. H., & Ong, J. (2018). The rationalization process of online game cheating behaviors. *Information, Communication & Society*, 21(2), 273-287.

Cheng, J., Danescu-Niculescu-Mizil, C., Leskovec, J., & Bernstein, M. (2017). Anyone can become a troll: analysis and simulation of online discussion sections show circumstances that can cause civil commentators to engage in aggressive behavior. *American Scientist*, 105(3), 152-156.

Croft, D. P., Madden, J. R., Franks, D. W., & James, R. (2011). Hypothesis testing in animal social networks. *Trends in ecology & evolution*, 26(10), 502-507.

Dimant, E. (2019). Contagion of pro-and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, 73, 66-88.

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological science*, 20(3), 393-398.

Gould, E. D., & Kaplan, T. R. (2011). Learning unethical practices from a co-worker: the peer effect of Jose Canseco. *Labour Economics*, 18(3), 338-348.

Han, Y., Ma, J., Bang, E., & Song, J. (2019). Dynamics of bullies and victims among Korean youth: A propensity score stratification analysis. *Children and Youth Services Review*, 98, 252-260.

Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics reports*, 519(3), 97-125.

Irdeto. (2018). Irdeto Global Gaming Survey. *Irdeto.com*. Retrieved from: https://resources.irdeto.com/irdeto-global-gaming-survey/irdeto-global-gaming-survey-report-2.

Jung, J., Busching, R., & Krahé, B. (2019). Catching aggression from one's peers: A longitudinal and multilevel analysis. *Social and personality psychology compass*, 13(2), e12433.

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science*, 322(5908), 1681-1685.

Kovanen, L., Karsai, M., Kaski, K., Kertész, J., & Saramäki, J. (2011). Temporal motifs in time-dependent networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(11), P11005.

Kwon, K. H., & Gruzd, A. (2017). Is offensive commenting contagious online? Examining public vs interpersonal swearing in response to Donald Trump's YouTube campaign videos. *Internet Research*, 27(4), 991-1010.

Lee, S. H., Kim, P. J., & Jeong, H. (2006). Statistical properties of sampled networks. *Physical review E*, 73(1), 016102.

Lofgren, E. T., & Fefferman, N. H. (2007). The untapped potential of virtual game worlds to shed light on real world epidemics. *The Lancet infectious diseases*, 7(9), 625-629.

Maiberg, E. (2014, April 30) Hacks! An investigation into the million-dollar business of video game cheating. *PCgamer.com*. Retrieved from: https://www.pcgamer.com/hacks-an-investigation-into-aimbot-dealers-wallhack-users-and-the-million-dollar-business-of-video-game-cheating/.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594), 824-827.

Murrell, A. R., Christoff, K. A., & Henning, K. R. (2007). Characteristics of domestic violence offenders: Associations with childhood exposure to violence. *Journal of family violence*, 22(7), 523-532.

Paranjape, A., Benson, A. R., & Leskovec, J. (2017, February). Motifs in temporal networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (pp. 601-610). ACM.

Park, J. K., Han, M. L., & Kim, H. K. (2015). A Study of Cheater Detection in FPS Game by using User Log Analysis. *Journal of Korea Game Society*, 15(3), 177-188.

PUBG. (2019, February 26) A Letter from the Anti-Cheat Team. *pubg.com*. Retrieved from: https://www.pubg.com/2019/02/26/a-letter-from-the-anti-cheat-team/.

Sampson, A. (2018, July 2) How We Discovered a Virus Infecting Tens of Thousands of Fortnite Players. *Rainway.com*. Retrieved from: https://rainway.com/blog/2018/07/02/how-we-discovered-a-virus-infecting-tens-of-thousands-of-fortnite-players/.

Srivastava, A. (2010). Motif analysis in the amazon product co-purchasing network. *arXiv* preprint arXiv:1012.4050.

Tsvetkova, M., García-Gavilanes, R., & Yasseri, T. (2016). Dynamics of disagreement: Large-scale temporal network analysis reveals negative interactions in online collaboration. *Scientific reports*, 6, 36333.

Tsvetkova, M., & Macy, M. W. (2015). The Social Contagion of Antisocial Behavior. *Sociological Science*, 2, 36-49.

Wu, Y., & Chen, V. H. H. (2013). A social-cognitive approach to online game cheating. *Computers in Human Behavior*, 29(6), 2557-2567.

Woo, J., Kang, S. W., Kim, H. K., & Park, J. (2018). Contagion of cheating behaviors in online social networks. *IEEE Access*, 6, 29098-29108.

Zuo, X., Gandy, C., Skvoretz, J., & Iamnitchi, A. (2016, March). Bad apples spoil the fun: Quantifying cheating in online gaming. In *Tenth International AAAI Conference on Web and Social Media*.

Appendix: Figures

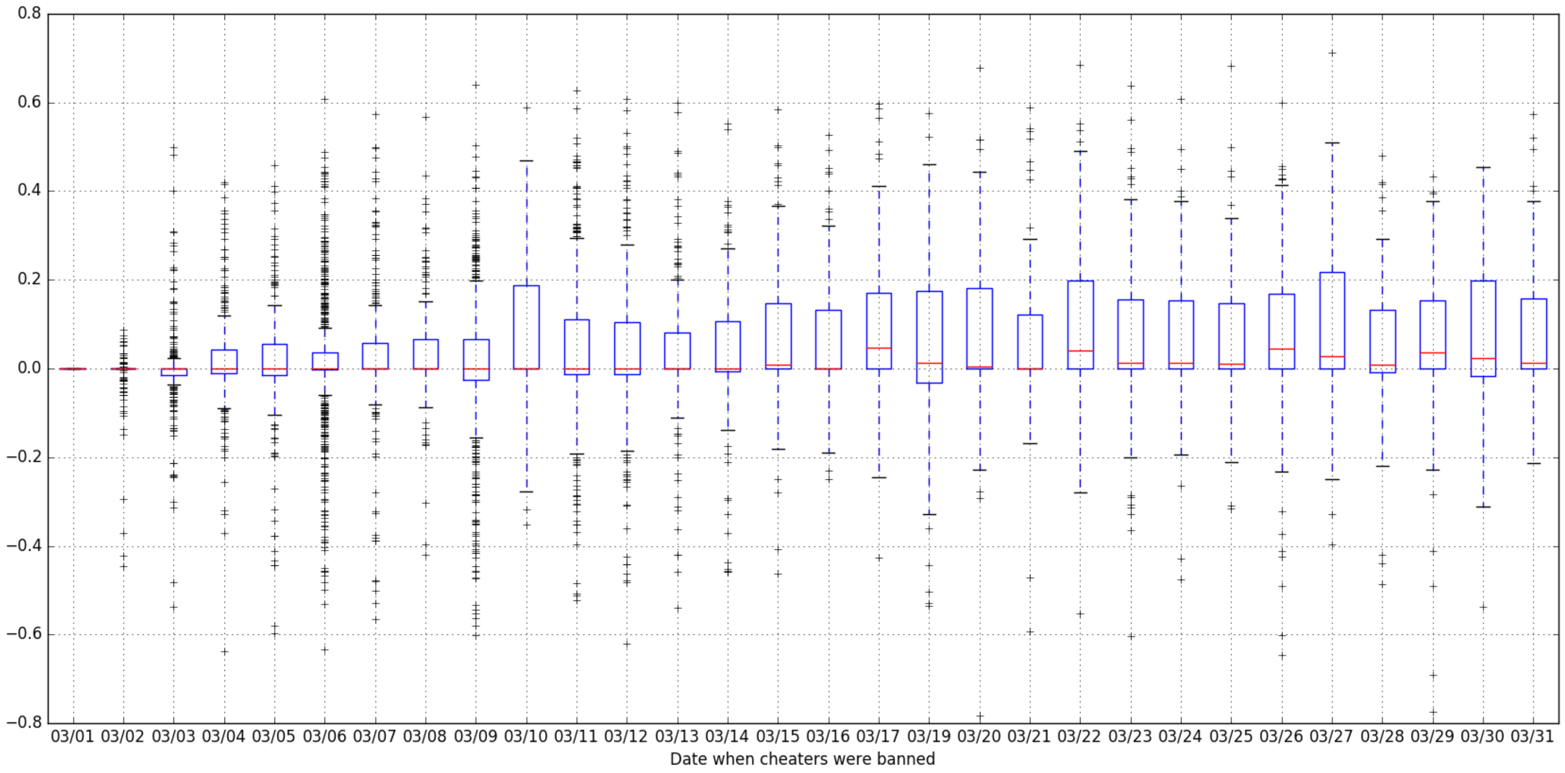

Fig. 1. Distribution of the difference between the recent kill ratio and average kill ratio of cheaters during the observation period

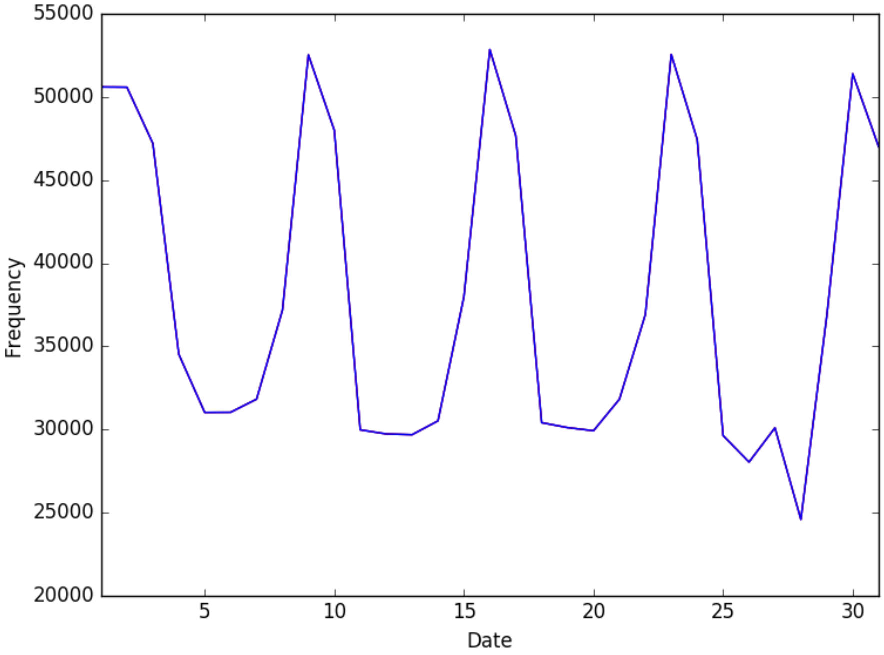

Fig. 2. Number of matches played by date

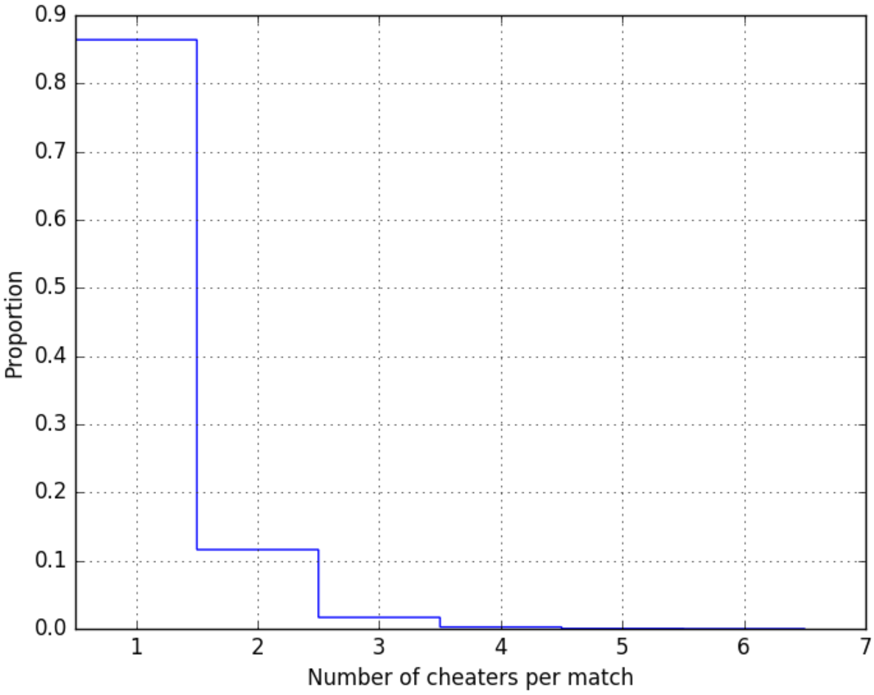

Fig. 3. Proportion of the number of cheaters per match (only matches with at least one cheater included)