

Cheating in Online Games: A Social Network Perspective

JEREMY BLACKBURN, University of South Florida
 NICOLAS KOURTELLIS, University of South Florida
 JOHN SKVORETZ, University of South Florida
 MATEI RIPEANU, University of British Columbia
 ADRIANA IAMNITCHI, University of South Florida

Online gaming is a multi-billion dollar industry that entertains a large, global population. One unfortunate phenomenon, however, poisons the competition and spoils the fun: cheating. The costs of cheating span from industry-supported expenditures to detect and limit it, to victims' monetary losses due to cyber crime.

This paper studies cheaters in the Steam Community, an online social network built on top of the world's dominant digital game delivery platform. We collected information about more than 12 million gamers connected in a global social network, of which more than 700 thousand have their profiles flagged as cheaters. We also observed timing information of the cheater flags, as well as the dynamics of the cheaters' social neighborhoods.

We discovered that cheaters are well embedded in the social and interaction networks: their network position is largely indistinguishable from that of fair players. Moreover, we noticed that the number of cheaters is not correlated with the geographical, real-world population density, or with the local popularity of the Steam Community. Also, we observed a social penalty involved with being labeled as a cheater: cheaters lose friends immediately after the cheating label is publicly applied.

Most importantly, we observed that cheating behavior spreads through a social mechanism: the number of cheater friends of a fair player is correlated with the likelihood of her becoming a cheater in the future. This allows us to propose ideas for limiting cheating contagion.

Categories and Subject Descriptors: H.3.4 [Information Storage and Retrieval]: Systems and Software—*Information networks*; K.4.1 [Computers and Society]: Public Policy Issues—*Ethics*; J.4 [Computer Applications]: Social and Behavioral Sciences—*Sociology*

General Terms: Measurement, Experimentation

Additional Key Words and Phrases: Social networks, Cheating in online games, Diffusion, Contagion

ACM Reference Format:

Jeremy Blackburn, Nicolas Kourtellis, John Skvoretz, Matei Ripeanu, and Adriana Iamnitchi, 2013. Cheating in Online Games: A Social Network Perspective ACM V, N, Article A (January YYYY), 24 pages. DOI = 10.1145/0000000.0000000 <http://doi.acm.org/10.1145/0000000.0000000>

1. INTRODUCTION

The popularity of online gaming supports a billion-dollar industry, but also a vigorous cheat code development community that facilitates unethical in-game behavior. “Cheats” are software components that implement game rule violations, such as seeing through walls or

This work is supported by the National Science Foundation, under grants CNS 0952420 and CNS 0831785. Author's addresses: Jeremy Blackburn, Nicolas Kourtellis and Adriana Iamnitchi, Department of Computer Science and Engineering, University of South Florida; John Skvoretz, Department of Sociology, University of South Florida; Matei Ripeanu, Department of Electrical and Computer Engineering, University of British Columbia.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© YYYY ACM 0000-0000/YYYY/01-ARTA \$15.00

DOI 10.1145/0000000.0000000 <http://doi.acm.org/10.1145/0000000.0000000>

automatically targeting a moving character. It has been recently estimated that cheat code developers generate between \$15,000 and \$50,000 per month from one class of cheats for a particular game alone [APB Reloaded Dev Blog 2011]. For some cheaters, the motivation is monetary: virtual goods are worth real-world money on eBay, and online game economies provide a lucrative opportunity for cyber criminals [Keegan et al. 2010; Ku et al. 2007]. For others, the motivation is simply a competitive advantage and the desire to win [Nazir et al. 2010]. And finally, some cheaters simply want to have fun and advance to a higher level in the game without investing tedious effort [Consalvo 2007].

In all cultures, players resent the unethical behavior that breaks the rules of the game: “The rules of a game are absolutely binding [...] As soon as the rules are transgressed, the whole play-world collapses. The game is over [Huizinga 1950]”. Online gamers are no different, judging by anecdotal evidence, vitriolic comments against cheaters on gaming blogs, and the resources invested by game developers to contain and punish cheating (typically through play restrictions).

Cheating is seen by the game development and distribution industry as both a monetary and a public relations problem [Consalvo 2007] and, consequently, significant resources are invested to contain it. For example, *Steam*, the largest digital distribution channel for PC games, employs the Valve Anti-Cheat System (VAC) that detects cheats and marks the corresponding user’s profile with a permanent, publicly visible (regardless of privacy setting), red, “ban(s) on record”. Game servers can be configured to be VAC-secured and reject players with a VAC-ban on record matching the family of games that the server supports. The overwhelming majority of servers available in the Steam server browser as of October 2011 are VAC-secured. For example, out of the 4,234 Team Fortress 2 servers available on October 12, 2011, 4,200 were VAC-secured. Of the 34 servers that were not VAC-secured, 26 were owned and administrated by a competitive gaming league that operates its own anti-cheat system.

Like many gaming environments, Steam allows its members to declare social relationships and connect themselves to *Steam Community*, an online social network. This work reports on our analysis of the Steam Community social graph with a particular focus on the position of the cheaters in the network. To enable this study, we crawled the Steam Community and collected data for more than 12 million user accounts. We also performed several additional rounds of focused data collection including daily ban status observations for 9 million users and daily neighborhood observations for newly banned cheaters. Our analysis targets the position of cheaters in the network, evidences homophily between cheaters, explores the geo-social characteristics that might differentiate cheaters from fair players, highlights the social consequences of the publicly visible cheating flag, and provides evidence in favor of a contagion process for the diffusion of unethical behavior at large-scale.

The remainder of this paper is organized as follows. We motivate this work in Section 2. Our datasets are presented in Section 3 (our data collection methodology is detailed in the Appendix). A brief analysis of socio-gaming characteristics is presented in Section 4. Section 5 analyzes the position of cheaters in the network from the perspective of declared relationships and the strength of their relationships measured via social-geographical metrics. It also presents the effect of the VAC-ban on individual players. Section 6 reasons about mechanisms for spreading the cheating behavior and provides an analysis of high-resolution data in support of a contagion mechanism. An overview of related work is presented in Section 7. Section 8 summarizes our findings and their consequences.

2. MOTIVATION

Multiplayer games break historical records for entertainment sales year after year with millions of players spending untold hours and billions of dollars [Cross 2011], pushing cutting-edge consumer hardware [Giles 2010], and spawning professional-level international “eSports” leagues with high-viewership live events and millions of dollars in prize money [Valve

2012a]. At the same time, gaming is increasingly relevant for sociologists and psychologists, as gaming interactions mimic, to some extent, real-world interactions [Szell and Thurner 2010], and the ground truth digital recording provides a large, precise dataset impossible to collect from laboratory experiments or surveys.

Dishonesty, prevalent in society [Ariely 2012], permeates the online gaming world as well, raising technical, legal and social challenges. Understanding cheaters' position in the social network that connects online gamers is relevant not only for evaluating and reasoning about anti-cheat actions and policies in gaming environments, but also for studying dishonest behavior in society at large.

For example, a study of large-scale cheating behavior can provide evidence to validate theories from social sciences and psychology on the nature of unethical behavior in general [Gino et al. 2009]. Studying a gaming network is particularly interesting because of the *competitive* nature of many multiplayer games, that has parallels in the real world, possibly describing corruption mechanisms in cases such as Enron, where "internal [group] competition could set the stage for the diffusion of 'widespread unethical behavior'" [Kulik et al. 2008].

Another prevalent cheating phenomenon is in academia, from plagiarizing in student assignments to falsifying data in research. Research into academic cheating has indicated the presence of a network effect. For example, the acceptance of a single high school cheater into a United States military service academy (typically thought of as bastions of honor and integrity) has been shown to cause a statistically significant 0.37 to 0.47 additional students to cheat [Carrell et al. 2008]. A study of 158 private universities [Rettinger and Kramer 2009] showed that observing other undergraduate students cheat was strongly correlated with one's own cheating behavior. What has been lacking, for the most part, is an empirical investigation into how cheating behavior diffuses through the relationships of a social network.

Another area where understanding cheating in games is relevant is gamification. Gamification attempts to import gameplay mechanics to otherwise non-gaming environments, with lofty goals such as increasing engagement, participation, and even performance. It has become a hot topic in both science [Thom et al. 2012; Werbach and Hunter 2012; Lin and Zhu 2012; Landers and Callan 2011; Deterding et al. 2011; Hamari and Lehdonvirta 2010] and industry [Foursquare Labs, Inc. 2012; Fitocracy, Inc. 2012; Gartner 2011], emphasizing even more the importance of understanding gaming, and the deviant behavior associated with it. As gamification becomes increasingly ubiquitous, the threat of cheating will become more prominent, and its effects more profound.

In fact, Foursquare, one of the most popular location-based social networks and an early innovator in gamification, has experienced "pervasive" cheating [Glas 2011]. In Foursquare, users "check-in" to physical locations, receiving points for various tasks. Badges are given to encourage service usage (e.g., the "Superstar" badge given for checking into 50 different venues), for various achievements (e.g., the "Douchebag" badge is given for checking in to 25 locations with a "douchebag" tag), and have been converted into a revenue stream by charging for sponsored "Partner Badges" which are often accompanied by special offers at a venue. Cheating not only diminishes the achievements of fair players, but poses a direct threat to Foursquare's business model as cheaters circumvent the rules set up by Foursquare's paying partners for badges and special offers.

Cheating in Foursquare is simple: lie about your location. By falsifying check-ins, a user can gain badges and honorifics that he would not attain legitimately. Effort has been invested into the detection of cheaters on Foursquare [Pelechrinis et al. 2012; He et al. 2011], but not to understanding how cheating propagates. Although useful, it seems unlikely that detection mechanisms will transfer to other gamified systems. While methods for protecting against and detecting cheaters tend to be domain specific, the process by which the cheating

behavior spreads is unlikely to differ in any fundamental sense, and thus remain applicable to the entire spectrum of gamified systems.

Finally, studying cheaters in online games can serve to better understand the behavior of individuals that abuse the shared social space in large-scale non-hierarchical communities. In online social networks, for example, such individuals abuse available, legal tools, like communication or tagging features, for marketing gains or political activism. Taken to the extreme, such behaviors lead to the tragedy of the commons: all game players become cheaters and then abandon the game, or corruption escalates and chaos ensues.

3. DATASETS

Steam controls between 50% and 70% of the PC game digital download market [Senior 2011], and claims over 40 million user accounts as of September 2012 [Valve 2012b]. Steam is run by Valve, who also develops some of the most successful multiplayer first-person shooter (FPS) games.

Games from a number of developers and publishers are available for purchase on Steam. A noteworthy segment is formed by the *multiplayer* FPS genre. In contrast to massively multiplayer online games, multiplayer FPSs usually take place in a relatively “small” environment, player actions generally do not affect the environment between sessions, and instead of one logical game world under the control of a single entity, there are often multiple individually-owned and operated servers. Because there is no central entity controlling game play and a large number of servers to choose from, the communities that form around individual servers are essential to the prolonged health of a particular game.

In our analysis we used several data sources. An initial dataset was collected by crawling the Steam Community website for user profiles and the social network represented by declared relationships between them. After a first round of analysis [Blackburn et al. 2012], we constructed a system that made daily observations of the ban status of a large set of users (over 9 million), as well as daily observations of the neighborhoods of newly banned cheaters and a set of control users.

3.1. The Steam Community

Steam Community is a social network comprised of Steam users, i.e., people who buy and play games on Steam. To have a Steam Community profile, one first needs to have a Steam account and take the additional step of configuring a profile. Users with a Steam account and no profile (and thus, not part of the Steam Community) can participate in all gaming activities, and can be befriended by other Steam users, but no direct information is available about them. Steam profiles are accessible in game via the Steam desktop and mobile clients, and are also available in a traditional web based format at <http://steamcommunity.com>.

Valve also provides the Valve Anti-Cheat (VAC) service that detects players who cheat and marks their profiles with a publicly visible, permanent *VAC ban*. Server operators can “VAC-secure” their servers: any player with a VAC ban for a given game can not play that game on VAC-secured servers (but they are allowed to play other games). In an effort to hinder the creators and distributors of cheats and hacks, most of the details of how VAC works are not made public. Valve describes VAC as sending periodic challenges to gamers’ machine that will, for example, execute an otherwise unused piece of game code and return a response [Kushner 2010]. If the player’s machine does not respond, then an alert of a possible cheat is registered. The detection itself operates similarly to anti-virus tools which examine changes in memory and signatures of known cheats. It is important to note that VAC is designed in a manner to leak as little information as possible to potential cheat creators: VAC is continuously updated and distributed piecemeal to obscure complete knowledge of the system. Additionally, VAC bans are not issued immediately upon cheat detection, but rather in delayed waves, as an additional attempt to slow an arms race between cheat creation and detection. Valve claims what amounts to a 0% false positive rate, and there are

Table I. The static snapshot of the *Steam Community* dataset as recorded in April 2011

Account	<i>Nodes</i>	<i>Edges</i>	<i>Profiles</i>	<i>Public</i>	<i>Private</i>	<i>Friends only</i>	<i>With location</i>
All users	12,479,765	88,557,725	10,191,296	9,025,656	313,710	851,930	4,681,829
Cheaters	-	-	720,469	628,025	46,270	46,174	312,354

only 10 known instances where bans were incorrectly handed out (and eventually lifted). It is difficult to ascertain VAC's false negative rate, but we believe that nearly all cheats are eventually detected.

While Steam accounts are free to create, they are severely restricted until associated with a verifiable identity, for example from game purchases (via a credit card) or from a gift from a verified account. *Once associated with an account, game licenses (whether bought or received as a gift) are non-transferable.* This serves as a disincentive for users to abandon flagged accounts for new ones: abandoning an account means abandoning all game licenses associated with that account. Moreover, Sybil attacks become infeasible, as they would require monetary investments and/or a real-world identity for even the most trivial actions, such as initiating chats with other players. Finally, with the introduction of a virtual goods trading platform, VAC bans in some games can now result in the “confiscation” of all virtual goods in an account that were not purchased with real money.

3.2. Static Snapshot

Although a fledgling Web API was available in early 2011, it did not expose a method for obtaining the friendslist of users. Thus, using the unmetered, consumable XML on the Steam Community web site, we crawled between March 16th and April 3rd, 2011. The crawler collected user profiles starting from a randomly generated set of SteamIDs and following the friendship relationships declared in user profiles. To seed our crawler, we generated 100,000 random SteamIDs within the key space, of which 6,445 matched configured profiles.

A Steam profile includes a nickname, a privacy setting (public, private, friends only or in-game only), set of friends (identified by *SteamIDs*), group memberships, list of games owned, the time spent playing each game for the past two weeks, life-time gameplay statistics, a user-selected geographical location, and a flag (VAC-ban) that indicates whether the corresponding user has been algorithmically found cheating. Although the cheating flag is publicly visible, Valve does not make the time the ban was applied available. The next section describes a second dataset that is augmented with ban dates.

From our initial 6,445 seeds of user IDs, we discovered just about 12.5 million user accounts, of which 10.2 million had a profile configured (about 9 million public, 313 thousand private, and 852 thousand visible to friends only). There are 88.5 million undirected *friendship* edges and 1.5 million user-created groups. Of the users with public profiles, 4.7 million had a location set (one of 33,333 pre-defined locations), 3.2 million users with public profiles played at least one game in the two weeks prior to our crawl, and 720 thousand users are flagged as cheaters. Table I gives the exact numbers.

3.3. Longitudinal Observations

The initial version of this work [Blackburn et al. 2012] used two static snapshots of the Steam social network as well as a 3rd party database of ban date observations (<http://www.vacbanned.com>) to perform temporal analysis. Unfortunately, the granularity of our observations was too coarse and the timestamps provided by the 3rd party database of unknown accuracy, which precluded trusted time analysis. In particular, we were interested in two data fields: first, the time when a VAC ban was applied. And second, the dynamics of the relationships of a newly branded cheater soon after the VAC ban was applied.

Fortunately, two new Web API methods were made available after our initial crawl. One method provides access to timestamped friends lists and another provides the ban states (without dates) of up to 100 users at a time. With these API methods, we made daily

Table II. Details of ban observations dataset

<i>Observations</i>	<i>Users</i>	<i>VAC bans</i>	<i>Comm. bans</i>	<i>Eco. probations</i>	<i>Eco. bans</i>
525,427,853	9,124,454	701,448	11,701	423	313

Table III. Details of neighborhood observations dataset

<i>Observations</i>	<i>Public control group</i>	<i>Public cheaters</i>	<i>Edges</i>	<i>Total Nodes</i>
349,123	8,712	2,337	294,058	284,765

observations of the ban status of users, as well as neighborhood observations of newly banned users and a control set of non-cheaters.

Ban status observations consist of the id of the observed user, the time stamp the observation was made, and a delta of the ban states. There are three different ban types: 1) **VACBanned**, 2) **CommunityBanned**, and 3) **EconomyBan**. We only make use of the VACBanned type in this work (the others are at early stages of being applied and there is no documentation describing their use).

Friends list observations consist of the time when the relationship was declared in addition to the ID of the friend. We note that deleted relationships are not recorded by Steam Community, which is why we record the neighborhood of selected gamers on a daily basis.

On June 20, 2012 we began daily observations of an initial set of 9,025,656 public profiles from the data set described in Section 3.1. (Details of the crawler implementation are presented in Section A.2) In total, we collected over 525 million ban status observations for over 9 million gamers, with Table II giving the exact numbers. We found an average of 83 gamers were flagged as cheaters per day, but, this number varied from 0 to over 400.

For monitoring the new cheaters' social neighborhood, any gamer that transitioned from a *VAC Banned = false* to *VAC Banned = true* state is treated specially. For these users, we begin a 10-day period of neighborhood observation where the friends list of the user is queried, and a delta stored once per day. Any friends of the user who do not already exist in the system are added, and will thus have their ban statuses recorded moving forward. In addition to users that transition from non-cheater to cheater, we also monitor the neighborhoods of a set of 10,000 randomly selected users (from the initial dataset) as a control group, 8,712 of which had public profiles. We call the combination of the control users and the newly discovered cheaters *monitored users*. We made 349,123 neighborhood observations of monitored users. Table III provides details on this dataset, and Appendix A describes the system we built to collect it.

4. CHEATERS AND THEIR GAMING HABITS

While the majority of this work is concerned with how cheaters are positioned within Steam Community, understanding their behavior as gamers helps to better understand their interactions in the community. To this end, we analyze the number of games owned and hours played per game genre using the tags provided by the Steam Store to describe each game and place games in the following categories: single-player, multi-player, single-player only, multi-player only, and co-op.

We use the categories “single-player” (the game can be played by a single human player) and “multi-player” (the game supports multiple human players) for two reasons: first, VAC bans only have an effect on multi-player games, and second, all games are tagged in at least one of these categories. Some games do not contain a single-player component at all. We classified these types of games as “multi-player only” if they were tagged as multi-player but not single-player, and those with no multi-player component are likewise classified as “single-player only”. Finally, “co-op”, or cooperative games, are loosely defined as multi-player games with a mechanic focusing on co-operative (as opposed to competitive) interaction between human players. For example, players might work together to defeat a horde of

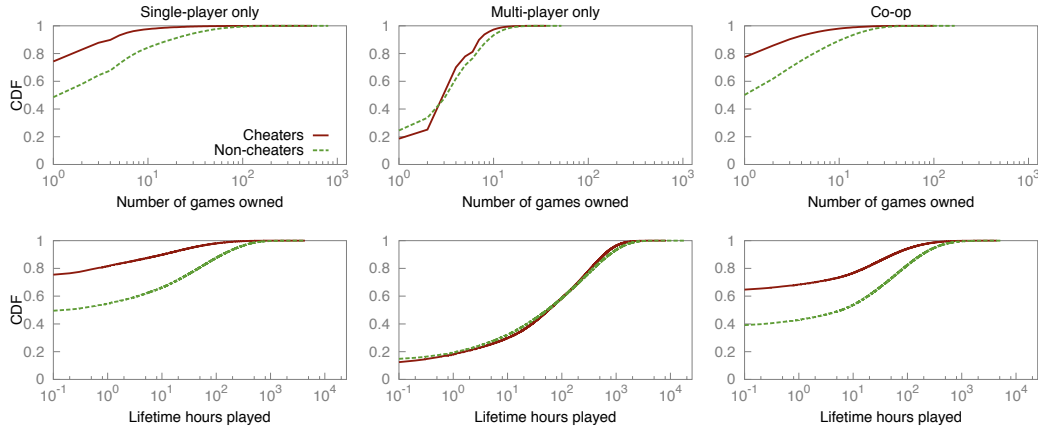


Fig. 1. The number of games owned and lifetime hours per genre for cheaters and non-cheaters from the March 2011 crawl, and newly discovered cheaters from the October 2011 crawl.

computer controlled goblins [Trendy Entertainment 2011], or to excavate a landscape and build a city [Re-logic 2011].

Figure 1 plots the *cumulative distribution function* (CDF) of the number of games owned and the lifetime hours on record per game category for cheaters and non-cheaters in our initial crawl (static snapshot). The CDF represents the probability of a random variable x having a value less than or equal to X . For example, around 60% of cheaters and non-cheaters have less than 100 lifetime hours played in multi-player only games. For brevity, we only present the plots representing the ownership of single-player only, multi-player only, and co-op games.

These results lead to the following observations. First, they provide confirmation of gaming as a social-activity: gamers on Steam Community are far more likely to own more than one multi-player games than single-player games, even though there are over twice as many single-player games available on Steam. This trend is even clearer when considering single-player only games vs. multi-player only games.

Next, we observe that non-cheaters are more likely to own more games than cheaters in general. However, the difference in number of games owned between cheaters and non-cheaters is significantly smaller for multi-player only games than for single-player only games. This provides an initial indication that cheaters are social gamers: even though they might not own as many games as a whole, they are as interested in multi-player games as non-cheaters are.

When considering the lifetime hours played per category, we see a similar story. Cheaters tend to play fewer hours of single-play only and co-op games than fair players. This is not entirely expected, as cheating can be motivated not only by competition with other players, but also for advancing in the game to access higher levels of fun [Dumitrica 2011]. Because the co-op tag was later introduced, it is possible that not all co-op games are properly tagged: however, of the games tagged as co-op, cheaters tend to own fewer games and play fewer hours than non-cheaters.

The message of this analysis is that cheaters are most definitely social gamers: they favor multi-player games over single-player games for both purchase and play time. Specifically, cheaters are much less interested in games *without* a multi-player component.

5. CHEATERS AND THEIR FRIENDS

One line of thought in moral philosophy is that (un)ethical behavior of an individual is heavily influenced by his social ties [Parfit 1984]. Under this theory, cheaters should appear

tightly connected to other cheaters in the social network. On the other hand, unlike in crime gangs [Ahmad et al. 2011], cheaters do not need to cooperate with each other to become more effective. Moreover, playing against other cheaters may not be particularly attractive as any advantage gained from cheating would be canceled out. These observations suggest that cheaters may be dispersed in the network, contradicting the first intuition.

To understand the position of cheaters in the social network, we characterize the Steam Community social network over three axes: First, we explore the relationship between cheating status and a user's number of friends (Section 5.1); second, we try to understand whether cheaters are visibly penalized by the other members of the social network (Section 5.2); and, finally, we explore the relationship between social network proximity and two other proximity metrics, geographical and community-based (Section 5.3).

5.1. Who is Friends with Cheaters?

Figure 2 presents the degree distribution for the Steam Community graph as a whole, for just cheater profiles, as well as for private, friends-only profiles, and for users without a profile, plotted as *complementary cumulative distribution functions* (CCDF). The CCDF represents the probability of a random variable x having a value greater than or equal to X^1 . For example, about 10% of Steam Community users have at least 20 friends. For users without a profile or with private profiles, edges in the graph are inferred based on the information from public profiles that declare the user as a friend. From the degree distributions we make two observations.

First, we discovered a hard limit of 250 friends (this limit has since been raised to 300 if a user links their Facebook account to their Steam Community account). However, there are some users who have managed to circumvent this hard limit. One user in particular has nearly 400 friends, and through manual examination we observed this user's degree increasing by one or two friends every few days. Coincidentally, this profile also has a VAC ban on record.

Second, all categories plotted in Figure 2, with the exception of that of users with Steam accounts but no profiles, overlap. This means that the distribution of the number of friends cheaters have is about the same as the non-cheaters' distribution. It also highlights that attempting to hide connection information through *private* or *friends-only* profile privacy settings is mostly unsuccessful: the player's position in the social network is revealed by the privacy settings of his friends.

While cheaters are mostly indistinguishable from fair players using the node degree distribution, a more important question is whether their behavior shows network effects. In other words, are cheaters more likely to be friends with other cheaters than with non-cheaters? Figure 3(a) plots the CDF of the fraction of a player's friends who are cheaters. Figure 3(b) plots the CCDF of the number of cheater friends for both cheaters and non-cheaters. This figure is comparable to Figure 2(a), but displays only the size of the cheating neighborhood.

The picture that emerges from these two figures is a striking amount of homophily between cheaters: cheaters are more likely to be friends with other cheaters. While nearly 70% of non-cheaters have *no* friends that are cheaters, 70% of cheaters have at least 10% cheaters as their friends. About 15% of cheaters have over half of their friends other cheaters.

While the differentiation is visually apparent, we ensured it was statistically significant via two methods: 1) the two sample Kolmogorov-Smirnov (KS) test, and 2) a permutation test to verify that the two samples are drawn from different probability distributions. An explanation of the two methods appears in Appendix B. We find that the distributions are in fact different with $p_{ks} < 0.01$, $D = 0.4367$, $p_{permute} < 0.01$, $T = 969.0140$, and $p_{ks} < 0.01$, $D = 0.4752$, $p_{permute} < 0.01$, $T = 766.8699$ for Figures 3(a) and 3(b), respectively.

¹Some definitions of the CCDF are strictly greater than, however, degree distributions in social networks are often plotted using the greater than or equal to definition.

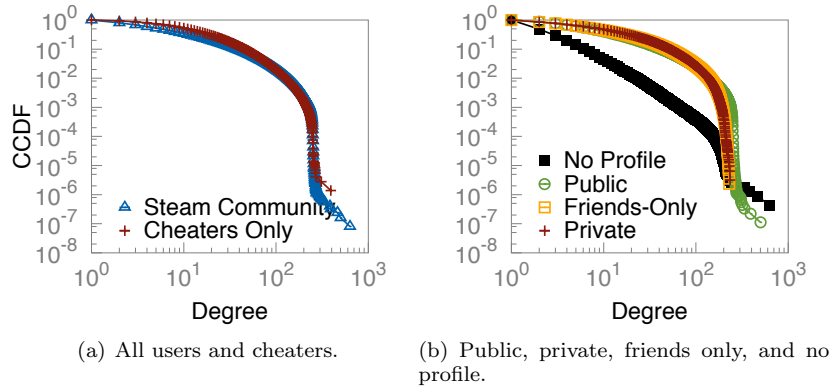


Fig. 2. Degree distributions in Steam Community

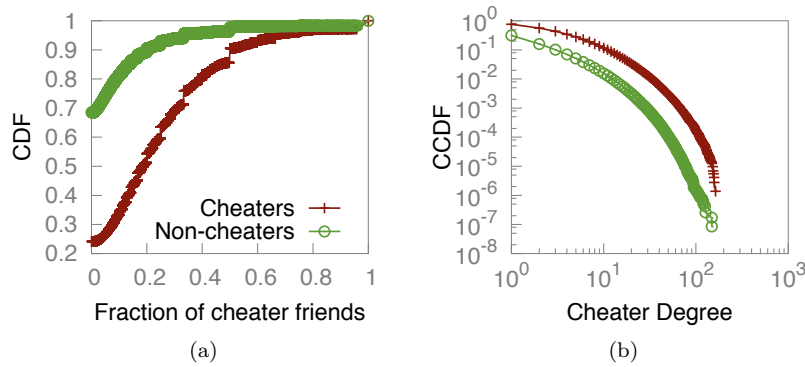


Fig. 3. (a) Fraction of cheaters' friends that are cheaters vs. the fraction of non-cheaters' friends that are cheaters. (b) CCDF of the number of cheater friends, the "cheater degree", for both cheaters and non-cheaters.

Declared relationships in online social networks are inconsistently backed by actual interactions [Wilson et al. 2009]. While gaming is a primarily interactive community, with the declared ties not always necessary for the primary purpose of the community, we verify that the number of declared friends translates into gaming interactions. To this end, Figures 4(a) and 4(b) plot the average number of games owned and hours played, respectively, in the two weeks prior to our March 2011 crawl as a function of social degree. There is a strong correlation between the degree of a user and the average number of hours played (0.70 for the whole Steam Community, 0.64 for cheaters). The figures show a positive correlation between the social degree of a gamer and the average number of games owned (0.38 Pearson correlation score) and a clear positive correlation for cheaters (0.63).

In other words, the investment in gaming (both monetary and time) increases with the number of friends for both cheaters and non-cheaters. Even though cheaters are involved in decidedly anti-social behavior, they still have a positive response to the social phenomena of gaming. This is an important result, as it introduces the possibility of VAC bans having more than just a utilitarian impact on cheaters: not only is their technical ability to game affected, but the ban might also impact their standing with their gameplay partners. In fact, in the next section we show that there are indeed negative social effects associated with being branded a cheater.

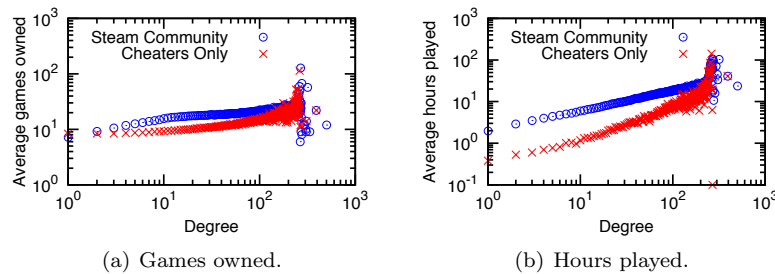


Fig. 4. The number of games owned, and hours played as function of the number of friends.

5.2. Are Cheaters Shunned?

While aggregate-level information shows some differentiation between cheaters and non-cheaters, the effect of the VAC-ban mark can better be understood by analyzing the reaction to the label. Using two static crawls 6 months apart, our previous work [Blackburn et al. 2012] found that newly banned cheaters were over twice as likely to change their profile to a more restrictive privacy state than non-cheaters. This indicates that cheaters themselves react to the cheating brand, perhaps in the naive hope that a more restrictive setting will provide a measure of protection from a potentially disapproving community.

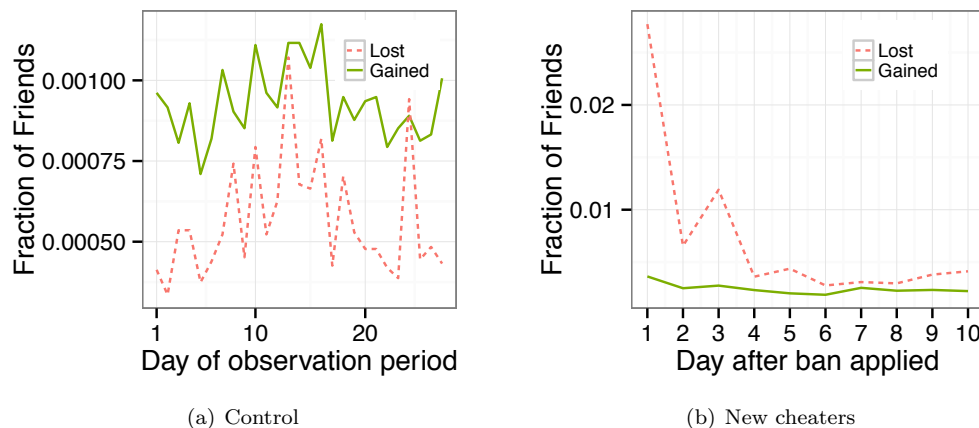


Fig. 5. The total fraction of friends gained and lost for the control group and new cheaters, relative to the number of friends on the first day of observation. Note different scales.

But is the community disapproving? Figure 5 plots the total fraction of friends lost and gained, relative to the first day of observation, for the control users and new cheaters from our longitudinal dataset described in Section 3.3. Confirming previous results [Blackburn et al. 2012], newly banned cheaters, while not completely ostracized, tend to lose many more friends than they gain.

From the high resolution dataset, we can now also observe that this reaction occurs very quickly: the greatest loss in friends happens within the first few days of the ban becoming publicly visible (Figure 5 (b)). While the control group’s pattern is what we might expect from a “clean up” of friends lists (where older, less active friendships are removed to make

room for new friends) as observed in [Xu et al. 2011], the speed and degree of friendship loss for the new cheaters indicates a response to the application of the cheating flag.

This new finding provides quantitative insight that human beings react to unethical behavior in a similar fashion as we do to epidemic disease, and indicates future directions towards modeling and suppressing the spread of unethical behavior. For example, the publicly visible cheating flag can be seen as a form of information-driven vaccination [Ruan et al. 2012]: knowledge of the unethical behavior is not hidden, and thus spreads through the network causing non-cheaters to sever ties with new cheaters. Models with both rewiring [Gross et al. 2006] and information-driven vaccination [Ruan et al. 2012] have been shown to have an effect on the spread of epidemic contagion in a network, and our analysis provides new empirical insight into these phenomena.

5.3. Are Cheaters Close?

We use two metrics, one geographical, and another social, to quantify the strength of the relationship among Steam Community users in general and among cheaters in particular.

5.3.1. Geographical Proximity. Exploring the relationship between geographical and social-network proximity may give quantitative support to the theory proposed in [Dumitrica 2011] according to which opinions on cheating are culturally derived. We thus first observe that user population on Steam Community does not follow real-world geographic population and, more importantly, that cheaters are not uniformly distributed with respect to geo-political boundaries.

Figure 6 shows the fraction of cheaters in the populations for the twelve countries comprising the union of the top ten user populations and the top ten cheater populations. The figure shows that cheaters are vastly overrepresented in some locations. For example, there are about 55,000 cheaters in the Nordic European countries (12.4% of the playing population of the region), while there are about 39,000 cheaters (3.9%) in the US. These regional differences become more striking when taking real-world population into account. In particular, we found enough Steam profiles to account for nearly 2.5% of Denmark’s 5.5 million residents, of which cheaters account for nearly 0.5% of Denmark’s population.

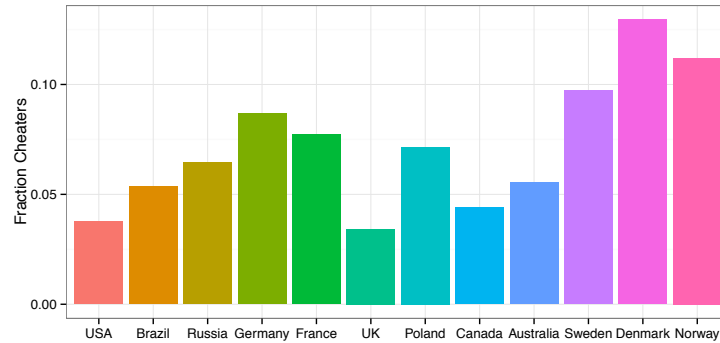


Fig. 6. The fraction of cheaters for the union of the top 10 player and top 10 cheater population countries. The countries are sorted on the x-axis in decreasing order of their real-world populations.

We now ask two additional questions: 1) Does the Steam Community exhibit properties of a location-based social network? and 2) Do cheaters tend to form geographically closer relationships with other cheaters than non-cheaters? To answer these, we measure node locality, a geo-social metric introduced in [Scellato et al. 2010]. The node locality of a given

node quantifies how close (geographically) it is to *all* of its neighbors in the social graph. Thus, a node locality of 1 indicates that a given node is at least as close to all of its neighbors as any other node in the graph is to their neighbors, and a value of 0 indicates that a given node is further away from all its neighbors than any other node in the graph.

We constructed the *location network* by including an edge from the social network if and only if both end points had a known location. Steam Community users can optionally choose to specify their location from a total of 33,333 possible locations at the country, state and city level. Because setting a locations is optional, this lead to a reduction in the size of the network, which, along with the geo-social properties of the resulting location network, can be seen in Table IV. We note that a subgraph composed solely of cheater-to-cheater relationships (C-C) has a lower mean inter-nodal distance and lower average link length than the location network as a whole.

Table IV. Location network properties: the number of nodes, edges, mean distance between users $\langle D_{uv} \rangle$, average link length $\langle l_{uv} \rangle$, average node locality $\langle NL \rangle$

Network	# of nodes	# of edges	$\langle D_{uv} \rangle$ (km)	$\langle l_{uv} \rangle$ (km)	$\langle NL \rangle$
Steam Community	4,342,670	26,475,896	5,896	1,853	0.79
Steam Community Cheater-to-Cheater	190,041	353,331	4,607	1,761	0.79
BrightKite	54,190	213,668	5,683	2,041	0.82
FourSquare	58,424	351,216	4,312	1,296	0.85

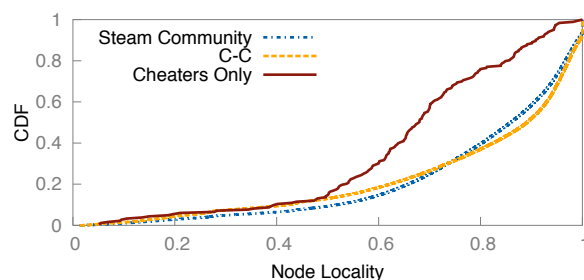


Fig. 7. CDF of node locality.

Figure 7 plots the CDF of node locality for the location network (Steam Community), the cheater-to-cheater subgraph (C-C), which includes only edges between cheaters, as well as just the cheaters within the location network (Cheaters Only), which includes all edges with at least one cheater. We first note that about 40% of users in the location network have a node locality of above 0.9, a phenomena exhibited by other geographic online social networks such as BrightKite and FourSquare [Scellato et al. 2010]. This is strong evidence that Steam Community relationships exhibit geo-social properties, a characteristic to be expected in the context of multiplayer gaming where high network latencies cannot be well masked by current game infrastructure. Next, we observe that the cheater-to-cheater network and the Steam Community at large have similar node locality distributions. Finally, when considering only the cheaters embedded within the location network, we see drastically lower node locality, with only about 10% of cheaters having a node locality greater than 0.9.

These results lead to three observations: 1) friendships tend to form between geographically close users, 2) cheaters tend to form relationships with other nearby cheaters and these links are geographically closer than those formed by non-cheaters, and 3) as evidenced by their lower node locality when considering the entire location network and not only the cheater-to-cheater subgraph, cheaters appear to befriend geographically remote

fair players. This might indicate that cheaters form relationships with other cheaters via a different mechanism than they form relationships with non-cheaters. Cheater-to-cheater relationships appear geographically constrained, while their relationships with non-cheaters are over larger distances.

Finally, we note that there are some caveats with these results. First, locations in Steam Community are set by the user, and to the best of our knowledge, not enforced in any way. In other words, cheaters might be lying about their locations. Second, the locations are selected from a predefined list, and users can specify only a country, or down to the city level. While these points should be taken into consideration, Steam Community clearly has properties of a location based social network, which is intuitive considering constraints on latency in multi-player games. In fact, in [Blackburn and Iamnitchi 2013], we studied 10 months of detailed server logs from a community owned and operated game server and discovered that interactions were dominated by declared friends. Since players choose servers based in large part on latency, this strengthens the premise that declared friends are geographically close. We thus believe that the effect of gamers potentially lying about their geographical location is inconsequential.

5.3.2. Social proximity. We use *social proximity* as the second metric to characterize the strength of the relationships between Steam Community users and understand whether they materially differ for the cheater population. The social proximity metric is based on a previous study [Onnela et al. 2007] that suggests that the overlap between the social neighborhood of two individuals is a good indicator of the strength of their relationship. We study the overlap of friends of users in the Steam Community networks to understand whether cheaters exhibit a stronger relationship with other cheaters than fair players do with fair players. We assess the strength of the relationship between two connected users by the overlap between their sets of friends, computed as follows:

$$Overlap_{uv} = m_{uv} / ((k_u - 1) + (k_v - 1) - m_{uv})$$

where m_{uv} is the number of common neighbors between users u and v , k_u is the number of neighbors of user u and k_v is the number of neighbors of user v . This overlap is calculated for two groups of user pairs: the 1.5 million pairs of cheaters (i.e., all cheater pairs in the full social network) and 1.5 million randomly selected pairs of non-cheaters (i.e., about 2% of the existing non-cheater pairs). Additionally, we also calculate the same metric on the cheater-only as well as on the non-cheater-only graphs.

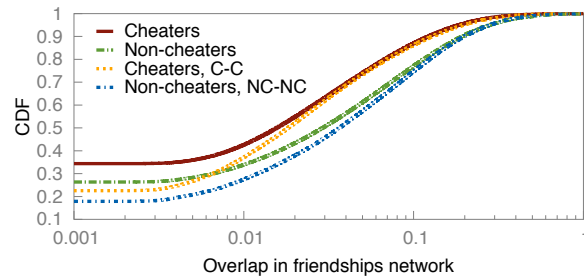


Fig. 8. CDF of social proximity for cheater and non-cheater pairs when we consider all relationships, only cheater to cheater relationships (labeled C-C) and only non-cheater to non-cheater relationships (labeled NC-NC).

Figure 8 shows a higher overlap for cheater pairs in the cheater-only graph and non-cheater pairs in the non-cheater-only graph compared to the respective overlaps in the

Table V. Percentage of cheaters found in top-N% of high degree centrality (DC) and betweenness centrality (BC) users in the Steam Community.

Top-N%	0.1	0.5	1.0	5.0	10.0
DC	3.25	4.46	5.11	7.06	8.20
BC	5.16	5.95	6.35	7.86	8.58

overall social network. This suggests that social relationships are weaker between different types of players (cheaters to non-cheaters) than within a uniform group.

6. PROPAGATION OF CHEATING

How does cheating behavior spread in the Steam Community? A first insight can be obtained by understanding whether cheaters hold positions of influence in the social network (Section 6.1). A higher-resolution understanding is reached by observing how cheating bans propagate in the network over time (Section 6.2).

6.1. Are Cheaters in Positions of Influence?

The position of a node in the social network affects the potential influence the node has on close or even remote parts of the network. For example, a high degree centrality node—one with many direct neighbors—can directly influence more nodes than a low degree centrality node. High betweenness centrality nodes, on the other hand, mediate the traffic along the shortest paths between many pairs of nodes, thus having influence on remote parts of the network. A high betweenness centrality cheater, for example, could facilitate the propagation of cheats and other deviant behavior to distant parts of the gamers network.

To understand the positional importance of the cheaters in the network, we study their potential for influence in the Steam Community by computing their degree centrality and node betweenness centrality. The degree centrality is simply the degree of the node in the network, and is thus a local metric. Betweenness centrality, however, is a global graph measure and consequently computationally expensive, requiring the calculation of the shortest paths between all pairs of nodes in the network. Due to the scale of our graph, we approximate betweenness centrality using κ -*path centrality*, a betweenness approximation method proposed in [Kourtellis et al. 2012].

We observe a high correlation of 0.9731 between degree and betweenness centrality scores of the gamers. This high correlation remains consistent when we differentiate on the player type: 0.9817 for cheaters, and 0.9726 for non-cheaters. Consequently, if a player has many friends in the Steam Community network, (that is, high degree centrality), not only can she influence many players directly, but she can also mediate the information flow between remote players due to her likely high betweenness centrality.

We focus only on the most central players in the network and study how many of them are cheaters. Table V demonstrates that cheaters are under-represented among the most central players, despite the fact that they have about the same degree distribution as the fair players, as shown earlier in Figure 2(a). Over 7% of the entire player population in our dataset are cheaters, but they make up less than 7% of the top 1% most central players, and are not adequately represented until we consider the top 5% to top 10% most central players. Earlier results from Section 5.2 provide an explanation for this. There seem to be social mechanisms that retard the growth of cheaters' social neighborhoods which could be preventing them from entering the top 1% central players in the social network. Surprisingly, this mechanism also limits their opportunity to influence remote parts of the network, as they tend not to be on central betweenness positions.

6.2. How Does Cheating Propagate over Time?

While remote influence is unlikely, is there any direct social influence in the propagation of cheating in the network? Figure 3 seems to suggest so, and thus we hypothesize that the friends of known cheaters are at risk of becoming cheaters themselves.

6.2.1. Number of Cheater Friends at the Time of the VAC Ban. We begin answering these questions by comparing the number of cheater friends for two groups of monitored users: non-cheaters and newly-labeled cheaters. Frequency distributions for the number of cheater friends for the monitored users (the number when banned for cheaters, and the number at the end of the observation period for control users) are plotted in Figures 9(a) and 9(b), respectively, with descriptive statistics listed in Table VI. Cheaters are *considerably* more likely to have multiple cheater friends at the time of the VAC ban application than our control group, and the samples are indeed drawn from different distributions ($D = 0.444$, $p_{ks} < 0.01$, $T = 36.80971$, $p_{perm} < 0.01$).

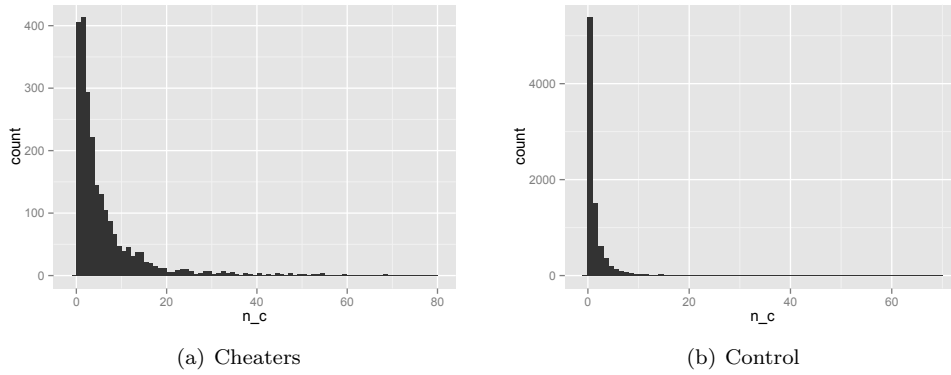


Fig. 9. Frequency distribution of the number of cheater friends that newly discovered cheaters and a control group of non-cheaters had as of Sept. 5, 2012. There are over three times as many cheaters with at least one cheater friend than there are with no cheater friends. In comparison, the number of control group users with at least 1 cheater friend is barely half the number with no cheater friends.

Table VI. Descriptive statistics for cheater friend distributions of the control group and new cheaters in the ban history dataset.

	n	mean	sd	median	min	max	se
Control	8,712	1.25	3.23	0.00	0.00	69.00	0.03
Cheaters	2,337	5.97	9.31	3.00	0.00	79.00	0.19

We built a logistic regression model using a single feature: the number of cheater friends at the time of the ban. Although there are many suitable classifiers, we use a logistic regression as it is a well understood method suitable for binary predictions. The classifier built from this high-resolution dataset achieved an area under the receiver operating characteristic curve (AUC) of 0.78, significantly higher than 0.61 which we achieved in the initial version of this work [Blackburn et al. 2012] where we used less accurate data. The AUC can be thought of the probability of the classifier ranking a cheater higher than a non-cheater for any random cheater/non-cheater pair. For our purposes, it shows that the number of cheater friends is a good predictor of becoming a cheater.

6.2.2. The Diffusion of Cheating Behavior. If the theory that cheating behavior is a social phenomenon that can be tracked via a corresponding social network holds true, then we should see newly discovered cheaters connected to previously discovered cheaters, and thus likely to form components when compared to the randomly selected control group.

We began by looking at the components formed by all monitored users and their social neighborhoods. The relative component size distributions for this network is plotted in Figure 10. The network had 6,692 components, with the largest containing 186,360 nodes and 200,236 edges. The existence of a large connected component forming from the 1-hop neighborhoods of $\approx 10,000$ users indicates that our monitored users are not randomly dispersed throughout Steam Community, and additional evidence that cheaters are indeed clustering together.

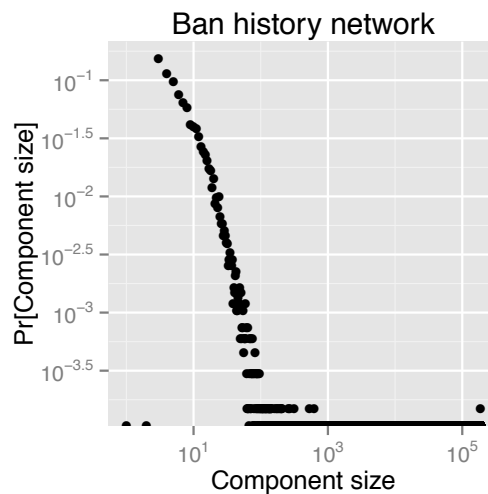


Fig. 10. Relative component size distribution for the ban history network. There are 6,692 components, and the largest component has 186,360 members.

Although the component distributions provide clear indication of cheaters clustering together, they do not necessarily address the issue of causality. To visualize the propagation of the cheating ban in the network (as a way of identifying causality), we plot in Figure 11 the progression of VAC bans in the 10 largest components composed entirely of monitored users between August 6th and September 5th 2012, divided into 6 intervals. We take network dynamics into account by coloring nodes and edges differently depending on the state of the network during the interval. Nodes are colored red if they were VAC banned by the end of the interval, and green otherwise. Edges are colored black if they existed prior to the end of the interval. Gray edges have not been formed by the end of the interval, but will exist at some point in the future. Finally, if an edge has been deleted (i.e., the friendship was dissolved) prior to the beginning of an interval it is colored orange.

Figure 11 shows evidence of spread. Most newly flagged cheaters are soon followed by their neighbors, resulting in clusters of cheaters appearing as time progresses. This discovery is the first indication of how rapid the contagion spreads through the network: out of the 279 cheaters at the end of Interval 6, only 28 (10%) were cheaters a month before.

A clarification is necessary at this point: we do not believe that the VAC ban is what inspires friends of newly-marked cheaters to cheat and thus become cheaters themselves; about 90% of the ties between eventual cheaters existed well before either were cheaters. Of

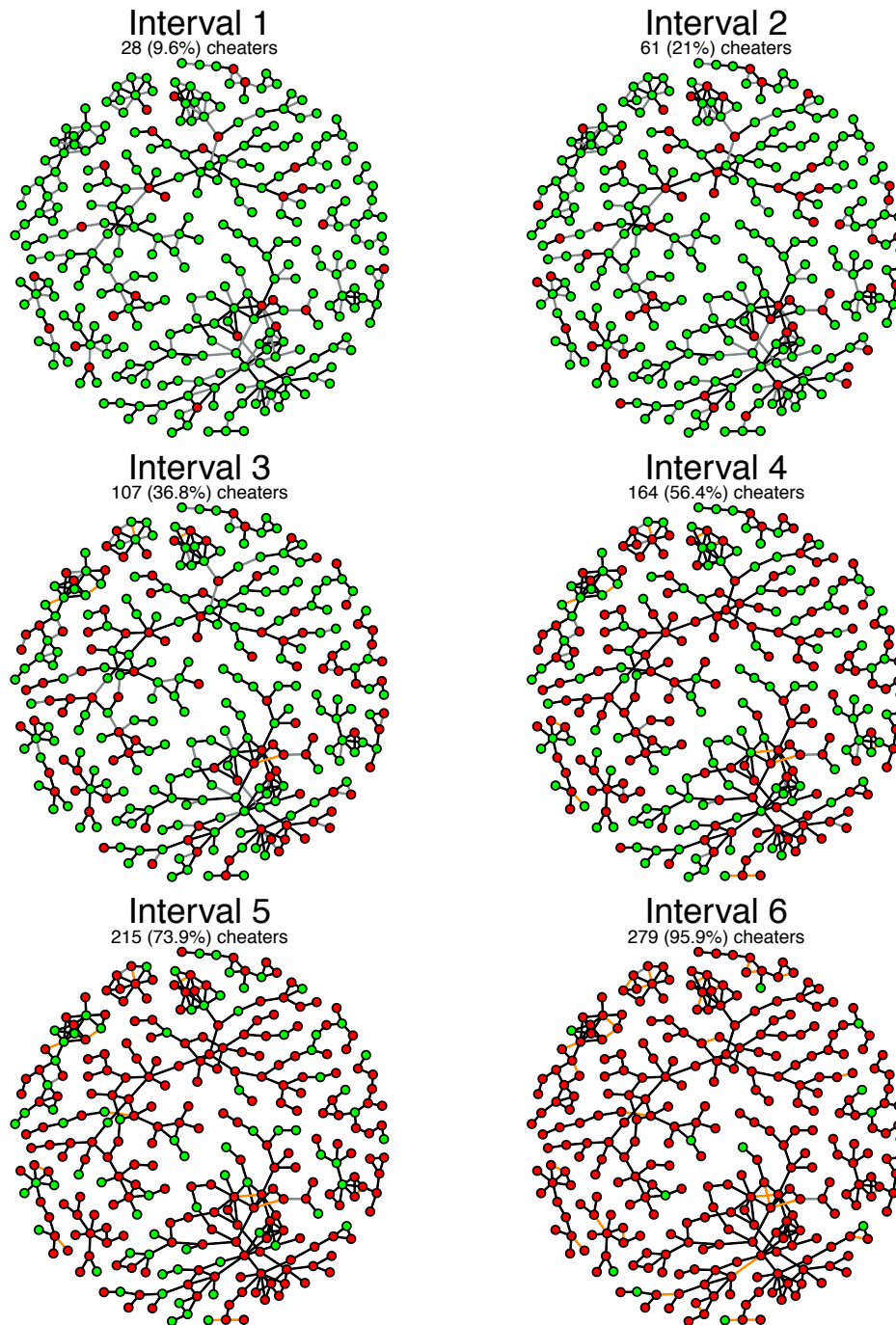


Fig. 11. The spread of cheating behavior in the 10 largest connected components of monitored users over a 1-month period, split into six intervals. If a non-cheater transitioned to cheater prior to the end of an interval, the node is colored red, and green otherwise. Gray edges will form in a future interval, black edges have formed by the end of an interval, and orange edges have been deleted prior to the beginning of an interval.

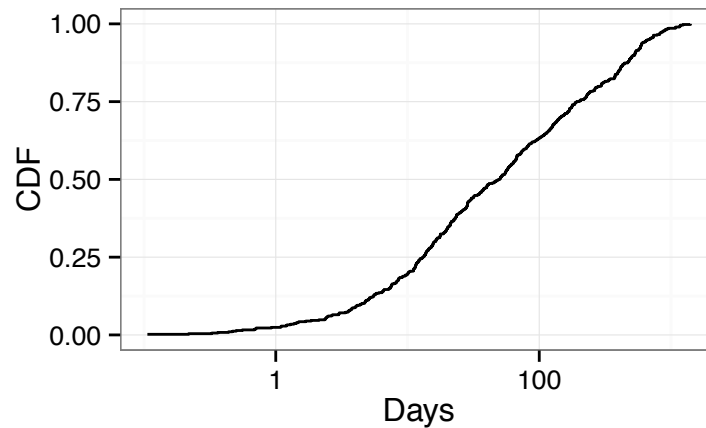


Fig. 12. CDF of the age of a relationship between pairs of eventual cheaters prior to either of the friends becoming a cheater. I.e., how old the relationship was before either friend cheated. 75% of the ties existed more than 10 days prior to the application of the cheating flag, about 50% had existed for more 50 days, and over 25% existed for more than 6 months. This indicates that the cheating behavior itself is not the catalyst for friendship formation, but rather that existing friendships allow the cheating behavior to spread.

these, 75% were created more than 10 days prior to one of the friends being labeled a cheater, as seen in Figure 12. In fact, the delayed application of the VAC ban obscures the exact timing of the contagion somewhat, and also allows friends to see the gains from cheating without the public penalty of the cheating label. Moreover, Figure 5 shows that non-cheaters sever their relationships with newly discovered cheaters, perhaps in an attempt to dissociate themselves from the dishonest behavior. What we believe happens is that the cheating behavior—implemented via sharing of cheat codes, for example—spreads along social ties prior to the cheating behavior being exhibited by the initial cheater in the relationship. When the VAC bans are applied some unknown time interval later, they reveal the time causality pattern of a social contagion.

There are two take-away lessons from this finding. First, it hints at possible improvements for the detection of cheaters. For example, the costs of detection efforts could be reduced by focusing on users who are in a relationship with a recently detected cheater. When used in conjunction with crowd-sourcing techniques to discover “spontaneous” cheaters (e.g., allowing users to report suspicious behavior), leveraging the structure of the social network would drastically reduce the set of users deemed at risk. The reduction in at-risk users would allow more computational expensive detection techniques to be deployed. This filtering would also make further use of crowd-sourcing possible, for example, by allowing other gamers to make judgments based on recorded gameplay sessions of at-risk users.

Second, knowledge of an epidemic process can also help *prevent* the adoption of cheating behavior in the first place. One solution could be to quarantine recently discovered cheaters (especially before the VAC ban becomes publicly visible), to limit the time of exposure to the “infectious” gamer. While an intuitive solution might be to remove cheaters from the service completely, this does not fit the business model of many gaming platforms. Even more, as we discovered in [Blackburn et al. 2012], cheaters are accepted by the communities of other games that they did not cheat in. For this reason, we suggest a *temporary* quarantine of new cheaters that prevents direct, out-of-game communication via chat for a limited time. This may retard the cheating contagion at practically no cost to the service provider by, for

example, allowing the delayed VAC ban to be applied before the newly discovered cheater is able to share his cheating tools via Steam Community's communication channels.

A more intriguing solution might be the development of interventions targeting at-risk users. In this case at-risk users could be actively targeted with a "vaccination" strategy, for example, personalized pop-up notifications reminding them of the consequences of cheating. Targeted interventions have been shown to effectively contain epidemic outbreaks, even with a lag between the identification of infected individuals and administration of the vaccine [Longini et al. 2005].

7. RELATED WORK

General gaming studies have characterized network traffic due to gaming, resource provisioning, work load prediction, and player churn in online games [Chambers et al. 2010; Claypool et al. 2003; Feng et al. 2002; Feng et al. 2005; Feng and Feng 2003]. Other studies have focused on the psychological and social properties of gamers [Fritsch et al. 2006] and gaming communities [Balint et al. 2011; Wei et al. 2010; Ducheneaut et al. 2007].

The social aspects of cheating in games, however, are a relatively unexplored. Nazir et al. study fake profiles created to gain an advantage in [Nazir et al. 2010]. Though the evaluation of behavior of player accounts within the social game *Fighters' Club* (FC), they are able to predict with high accuracy whether a profile is fake. Users in FC cheat by performing what is essentially a Sybil attack, since a player's power in FC is directly proportional to the number of declared friends that are also players. Cheaters in Steam Community, however, do not explicitly gain an advantage by altering the structure of the social graph. Instead, they are attacking out of band game rule implementations.

"Gold farmers" are cheaters that make black-market exchanges of real world currency for virtual goods outside of sanctioned, in-game, trade mechanisms. By examining social networks constructed from database dumps of banned *EverQuest II* (EQ2) players, Keegan et al. [Keegan et al. 2010] found gold farmers exhibit different connectivity and assortativity than both their intermediaries and normal players, and are similar to real-world drug trafficking networks. Ahmad et al. [Ahmad et al. 2011] further examined trade networks of gold farmers and propose models for deviant behavior prediction.

Their dataset differs from ours in both motivation for cheating, and the method of punishing cheaters. No clear financial motivation for cheating exists in the majority of games played by the majority of Steam Community players. Although the growing eSports industry has made gaming a possible profession, and tournaments with real world prizes are seen as a stepping stone for amateurs to go professional, most gameplay time is not spent with prize money on the line. While trafficking of virtual goods most certainly has implications for the monetization of the Free-to-Play model that has been recently adopted by many games, VAC bans apply specifically to *cheating*. Additionally, while cheaters in EQ2 have their accounts permanently disabled, cheaters in Steam Community are only restricted from playing the particular game they were caught cheating in on VAC-secured servers, as explained in Section 3.

Finally, we note that to the best of our knowledge, our work is the largest scale study of cheaters in a gaming social network. We discovered over double the amount of cheaters as there were players in [Nazir et al. 2010], and multiple orders of magnitude more cheaters than players in [Keegan et al. 2010; Ahmad et al. 2011].

Although not much quantitative analysis has been performed, social aspects of cheating in video games have been studied qualitatively. Duh and Chen [Duh and Chen 2009] describe several frameworks for analyzing cheating, as well as how different cheats can impact online communities. Dumitrica [Dumitrica 2011] examines *Neopets*, a web based social game. She describes a process by which gamers, who naturally seek ways to increase their "gaming capital", are tempted to cheat, and argues that a cheating culture emerges, where social values are used to understand and evaluate the ethical questions of cheating.

Although Steam Community in particular has not been studied before, other social networks of online gamers have been addressed in recent studies. Szell and Thurner [Szell and Thurner 2010] provide a detailed analysis of the social interactions between players in *Pardus*, a web-based Massively Multiplayer Online Game. By studying the socio-economic aspects of players, they provide additional evidence that communities of gamers can serve as a proxy for real world behavior. Xu et al. interviewed 14 Halo 3 players to study the meaning of relationships within an online gaming context [Xu et al. 2011]. They found evidence of in-game relationships supported by real-world relationships, triadic closure of relationships making use of both real and virtual relationships as a bridge, and in-game interactions strengthening ties in the real world. They further found evidence of social control as a tool for managing deviant behavior. Mason and Clauset investigated the behavior of Halo: Reach (the 2010 sequel to Halo 3) players by combining gameplay logs with psychometrics and a social network constructed from survey data [Mason and Clauset 2013]. They find that gamers preferred to play with friends, that teaming up with friends increased performance, and that social ties fundamentally affect the environment of competition.

Although these works did not study cheaters, their findings help make the case for the transferability of our work to the real world. In particular, Szell and Thurner’s findings that social structures formed by gamers mirror those of real life communities indicates that the patterns of cheating diffusion we observed likely take place in domains outside of gaming. Further, the fact that gaming activity can be used as a predictor for meaningful relationships as discovered by Xu et al. and Mason and Clauset, mitigates concerns that Steam Community relationships might not properly model social relationships; an issue for studies of OSNs like Facebook [Wilson et al. 2009].

The diffusion of (un)ethical behavior in social networks has been a topic of interest in business and management in particular. Kulik et al. proposed a theory that competitive environments can act as a catalyst for the spread of unethical behavior, using Enron as a case study [Kulik et al. 2008]. Enron’s internal corporate structure was built around the concept of “stacking”, where employees would be ranked on their performance and split along percentiles into different groups. Those in the top percentile group would receive commensurate benefits, while those in the lowest 15% were placed into an isolated area together. Those in the lower 15% were fired within a few weeks if they were unable to find a new job within Enron, which was a near certainty due to their publicly known poor performance. Ultimately, employees could be divided into two groups: winners and losers. This intense competition lead some employees to cheat: e.g., inflating the projected profit of a proposal. An honest employee could not compete, became a loser, and was summarily fired. After observing the gains made by cheating co-workers, Kulik, O’Fallon, and Salimanh argue the behavior would be emulated by otherwise ethical employees.

Similarly, the overwhelming majority of multiplayer games have clear winners and losers. In multiplayer FPSs like Halo and TF2, performance is also made public, as it was in Enron. At minimum, these games tend to have an in-game scoreboard displaying performance related statistics for each player (such as points, or the number of kills and deaths in the given gaming session), with the highest performing players at the top of the scoreboard’s visual representation, and the lowest at the bottom. In addition to the in-game scoreboard, most games make world-wide and friends leaderboards, as well as more detailed life-time statistics such as accuracy, number of games won, and a variety of game-specific statistics available.

Although no one is fired for poor performance in a game, there are consequences for poor performance besides the social stigma of being a “newb”. For example, the TrueSkill [Herbrich et al. 2007], algorithm used by Halo, ranks players and places similarly skilled individuals in the same game. The rank of a player is visible in the in-game UI for Halo 3 (studied by Xu et al. [Xu et al. 2011]), removed from the UI of its sequel Halo: Reach (studied by Mason and Clauset [Mason and Clauset 2013]), and will eventually be made available on the

stats tracking website for the latest incarnation, 2012's Halo 4², but it affects matchmaking none the less. A form of cheating known as "boosting" has sprung up where players boost their statistics using various unscrupulous methods. There is even at least one service that claims to employ professional gamers to boost accounts, later selling them for around \$50 a piece [Halotags 2012]. The competitive aspects shared by gaming and business indicates they share same process of unethical behavior's spread. Thus, findings from this work can help explain how unethical behavior spread through an organization like Enron.

8. SUMMARY AND DISCUSSION

Online gaming is poised to become the largest revenue-generating segment of the entertainment industry, with millions of geographically dispersed players engaging each other within the confines of virtual worlds. An ethical system is created along with the rules that govern the games. Just like in the real world, some players make the decision to circumvent the established rules to gain an unfair advantage, a practice actively discouraged by the industry and frowned upon by gamers themselves. This paper examined characteristics of these unethical actors in a large online gaming social network.

At a high level, viewed from the perspective of global network metrics, cheaters are well embedded in the social network, largely indistinguishable from fair players. This is not entirely unexpected. Cheaters are still gamers, and even though they are permanently marked, they remain members of the community. We observed evidence of this by examining socio-gaming metrics as well as the social network.

However, when we examine the transition from fair player to cheater, we observe the effects of the cheating brand. First, cheating behavior spreads via a contagion mechanism, where the presence and the number of cheater friends of a fair player is correlated with the likelihood of her becoming a cheater in the future. We observed clusters of cheaters forming as the "infection" of unethical behavior spread through the network. Consequently, cheaters end up having more cheater friends than the non-cheaters have. Second, we observed that cheaters are likely to switch to more restrictive privacy settings once they are caught, a sign that they might be uncomfortable with the VAC ban. We also found that cheaters lose friends over time compared to non-cheaters, an indication that there is a social penalty involved with cheating.

Cheater distribution does not follow geographical, real-world population density. The fact that some regions have higher percentages of cheaters to the player population suggests that cheating behavior may be related to differences between specific geo-social cultures. Such cheating-prone communities might be the target of more scrutiny, or the result of higher tolerance to cheating behavior, both in the real-world and gaming population.

Our study has consequences for gaming in particular, but also for other online social networks with unethical members. In the case of gaming, individual servers can evaluate the cheating risk of a new player by looking at a combination of attributes inferred from the player's profile that include structural features. In the case of general online social networks, the findings of our study can be used to better understand the effects of countermeasures to deal with anti-social behavior. For example, the profiles of users who abuse the available communication tools for political activism or personal marketing, or who appear to automate their actions could be publicly tagged. Our study gives a preliminary indication that, over time, the reaction of fair users to such information will make it harder to benefit from forms of anti-social behaviors that attempt to harness network effects. The fair users tend to have a vested interest in maintaining the quality of the shared social space and will limit the connectivity of the abusing profiles.

²The ranking system has become so important to players that developers are postponing the release until they can properly tune it, much to the chagrin of a sometimes rabid fan base hungry for a metric by which to gauge their standing in the highly competitive community.

Finally, we provided large scale empirical confirmation of theories regarding the (un)ethical behavior at large. The competitive environment of gaming in many ways mirrors that of the business world and cutthroat organizations where unethical behavior might run rampant, such as Enron. By harnessing the wealth of data online social networks provide, and the black-and-white nature of the VAC ban, we greatly reduce the uncertainty present in previous survey based studies. As unethical behavior, and cheating in particular, is an omnipresent threat in society, our study might serve as a basis for further understanding this seemingly counter-productive behavior. Ultimately, we believe that future analysis and modeling of the rich dataset generated by gamers will provide insight into not only the mechanism by which (un)ethical actions pervade a community, but might also illuminate methods for stanching the flow of societally destructive unethical behavior.

REFERENCES

- AHMAD, M., KEEGAN, B., SULLIVAN, S., WILLIAMS, D., SRIVASTAVA, J., AND CONTRACTOR, N. 2011. Illicit Bits: Detecting and Analyzing Contraband Networks in Massively Multiplayer Online Games. In *Proceedings of the IEEE third international conference on social computing. SocCom '11*. IEEE, 127–134.
- APB RELOADED DEV BLOG. 2011. <http://goo.gl/m9k1r>.
- ARIELY, D. 2012. *The (Honest) Truth About Dishonesty: How We Lie to Everyone—Especially Ourselves*. Harper.
- BALINT, M., POSEA, V., DIMITRIU, A., AND IOSUP, A. 2011. An analysis of social gaming networks in online and face to face bridge communities. In *Proceedings of the third international workshop on Large-scale system and application performance*. ACM, New York, NY, USA, 35–42.
- BLACKBURN, J. AND IAMNITCHI, A. 2013. Relationships under the Microscope with Interaction-Backed Social Networks. In *The first international conference on Internet Science 2013*.
- BLACKBURN, J., SIMHA, R., KOURTELLIS, N., ZUO, X., RIPEANU, M., SKVORETZ, J., AND IAMNITCHI, A. 2012. Branded with a scarlet "C": cheaters in a gaming social network. In *Proceedings of the 21st international conference on World Wide Web*. ACM, New York, NY, USA, 81–90.
- CARRELL, S., MALMSTROM, F., AND WEST, J. 2008. Peer effects in academic cheating. *Journal of human resources* 43, 1, 173–207.
- CHAMBERS, C., FENG, W.-C., SAHU, S., SAHA, D., AND BRANDT, D. 2010. Characterizing online games. *IEEE/ACM Transactions on Networking* 18, 3, 899–910.
- CLAYPOOL, M., LAPOINT, D., AND WINSLOW, J. 2003. Network analysis of Counter-strike and Starcraft. In *Performance, Computing, and Communications Conference, 2003. Conference Proceedings of the 2003 IEEE International*. 261–268.
- CONSALVO, M. 2007. *Cheating: Gaining Advantage in Videogames* 1st, Ed. MIT Press, Cambridge, Massachusetts.
- CROSS, T. 2011. All the world's a game. <http://www.economist.com/node/21541164>.
- DETERDING, S., SICART, M., NACKE, L., O'HARA, K., AND DIXON, D. 2011. Gamification. using game-design elements in non-gaming contexts. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '11. ACM, New York, NY, USA, 2425–2428.
- DUCHENEAUT, N., YEE, N., NICKELL, E., AND MOORE, R. J. 2007. The life and death of online gaming communities: a look at guilds in world of warcraft. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM.
- DUH, H. B.-L. AND CHEN, V. H. H. 2009. Cheating Behaviors in Online Gaming. *Online Communities and Social Computing* 5621, 1–7.
- DUMITRICA, D. D. 2011. An exploration of cheating in a virtual gaming world. *Journal of Gaming & Virtual Worlds* 3, 1, 21–36.
- FENG, W., CHANG, F., FENG, W., AND WALPOLE, J. 2005. A traffic characterization of popular on-line games. *IEEE/ACM TON* 13, 488–500.
- FENG, W.-C., CHANG, F., FENG, W.-C., AND WALPOLE, J. 2002. Provisioning on-line games: a traffic analysis of a busy counter-strike server. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurement*. ACM, New York, NY, USA, 151–156.
- FENG, W.-C. AND FENG, W.-C. 2003. On the Geographic Distribution of On-line Game Servers and Players. In *the 2nd workshop*. ACM Press, New York, New York, USA, 173–179.
- FITOCRACY, INC. 2012. Fitocracy. <http://www.fitocracy.com>.
- FOURSQUARE LABS, INC. 2012. Foursquare. <http://www.foursquare.com>.

- FRITSCH, T., VOIGT, B., AND SCHILLER, J. 2006. Distribution of online hardcore player behavior: (how hardcore are you?). In *Proceedings of 5th ACM SIGCOMM workshop on Network and system support for games*. ACM, New York, NY, USA.
- GARTNER. 2011. Gartner says by 2015, more than 50 percent of organizations that manage innovation processes will gamify those processes. <http://www.gartner.com/newsroom/id/1629214>.
- GILES, J. 2010. Inside the race to hack the kinect. *New Scientist* 208, 2789, 22 – 23.
- GINO, F., AYAL, S., AND ARIELY, D. 2009. Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel. *Psychological Science* 20, 3, 393–398.
- GLAS, R. 2011. Breaking reality: exploring pervasive cheating in Foursquare. *Proceedings of DiGRA 2011*, 1–15.
- GROSS, T., D’LIMA, C. J. D., AND BLASIUS, B. 2006. Epidemic Dynamics on an Adaptive Network. *Phys. Rev. Lett.* 96, 208701.
- HALOTAGS. 2012. Halotags halo 3 & halo 4 mlg level 50 accounts for sale. <http://www.halotags.com>.
- HAMARI, J. AND LEHDONVIRTA, V. 2010. Game design as marketing: How game mechanics create demand for virtual goods. *International Journal of Business Science & Applied Management* 5, 1, 14–29.
- HE, W., LIU, X., AND REN, M. 2011. Location cheating: A security challenge to location-based social network services. In *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*. IEEE, 740–749.
- HERBRICH, R., MINKA, T., AND GRAEPEL, T. 2007. Trueskill: A bayesian skill rating system. *Advances in Neural Information Processing Systems* 19, 569.
- HOTHORN, T., HORNIK, K., VAN DE WIEL, M. A., AND ZEILEIS, A. 2008. Implementing a class of permutation tests: The coin package. *Journal of Statistical Software* 28, 8, 1–23.
- HUIZINGA, J. 1950. *Home Ludens: A Study of the Play-Element in Culture* reprint Ed. Beacon Press, Boston, Massachusetts.
- KEEGAN, B., AHMED, M., WILLIAMS, D., SRIVASTAVA, J., AND CONTRACTOR, N. 2010. Dark Gold: Statistical Properties of Clandestine Networks in Massively Multiplayer Online Games. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*. 201–208.
- KOURTELLIS, N., ALAHAKOON, T., SIMHA, R., IAMNITCHI, A., AND TRIPATHI, R. 2012. Identifying high betweenness centrality nodes in large social networks. *Social Network Analysis and Mining*, 1–16.
- KU, Y., CHEN, Y., WU, K., AND CHIU, C. 2007. An empirical analysis of online gaming crime characteristics from 2002 to 2004. In *PAISI 07*. 34–45.
- KULIK, B. W., O’FALLON, M. J., AND SALIMATH, M. S. 2008. Do Competitive Environments Lead to the Rise and Spread of Unethical Behavior? Parallels from Enron. *Journal of Business Ethics* 83, 4, 703–723.
- KUSHNER, D. 2010. Steamed: Valve software battles video-game cheaters. <http://goo.gl/wWgq6n>.
- LANDERS, R. AND CALLAN, R. 2011. Casual social games as serious games: The psychology of gamification in undergraduate education and employee training. *Serious Games and Edutainment Applications*, 399–423.
- LIN, R. AND ZHU, X. 2012. Leveraging social media for preventive care-a gamification system and insights. *Studies in health technology and informatics* 180, 838.
- LONGINI, I. M., NIZAM, A., XU, S., UNGCHUSAK, K., HANSHAOWORAKUL, W., CUMMINGS, D. A. T., AND HALLORAN, M. E. 2005. Containing Pandemic Influenza at the Source. *Science* 309, 5737, 1083–1087.
- MASON, W. AND CLAUSET, A. 2013. Friends FTW! friendship and competition in halo: reach. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 375–386.
- NAZIR, A., RAZA, S., CHUAH, C.-N., AND SCHIPPER, B. 2010. Ghostbusting facebook: detecting and characterizing phantom profiles in online social gaming applications. *WOSN’10: Proceedings of the 3rd conference on Online social networks*.
- ONNELA, J.-P., SARAMÄKI, J., HYVÖNEN, J., SZABÓ, G., LAZER, D., KASKI, K., KERTÉSZ, J., AND BARABÁSI, A.-L. 2007. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences* 104, 18, 7332–7336.
- PARFIT, D. 1984. *Reasons and Persons*. Oxford University Press.
- PELECHRINIS, K., KRISHNAMURTHY, P., AND ZHANG, K. 2012. Gaming the Game: Honeypot Venues Against Cheaters in Location-based Social Networks. *arXiv*.
- RE-LOGIC. 2011. Terraria. <http://store.steampowered.com/app/105600/>.
- RETTINGER, D. AND KRAMER, Y. 2009. Situational and Personal Causes of Student Cheating. *Research in Higher Education* 50, 293–313.
- RUAN, Z., TANG, M., AND LIU, Z. 2012. Epidemic spreading with information-driven vaccination. *Phys. Rev. E* 86, 036117.

- SCCELLATO, S., MASCOLO, C., MUSOLESI, M., AND LATORA, V. 2010. Distance matters: geo-social metrics for online social networks. *WOSN'10: Proceedings of the 3rd conference on Online social networks*.
- SEKHON, J. S. 2011. Multivariate and propensity score matching software with automated balance optimization: The matching package for r. *Journal of Statistical Software* 42, 7, 1–52.
- SENIOR, T. 2011. Steam controls up to 70% of the pc download market and is “tremendously profitable”. <http://goo.gl/MjgA3>.
- STAUDT, S. 2011. Steam condenser. <https://github.com/koraktor/steam-condenser>.
- SZELL, M. AND THURNER, S. 2010. Measuring social dynamics in a massive multiplayer online game. *Social Networks* 32, 313–329.
- THOM, J., MILLEN, D., AND DiMICCO, J. 2012. Removing gamification from an enterprise sns. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. CSCW '12. ACM, New York, NY, USA, 1067–1070.
- TRENDY ENTERTAINMENT. 2011. Dungeon Defenders. <http://store.steampowered.com/app/65800/>.
- VALVE. 2012a. The international - dota 2. <http://www.dota2.com/international>.
- VALVE. 2012b. Steam community update now live. <http://store.steampowered.com/news/8788/>.
- WEI, X., YANG, J., ADAMIC, L., ARAÚJO, R., AND REKHI, M. 2010. Diffusion dynamics of games on online social networks. *WOSN'10: Proceedings of the 3rd conference on Online social networks*.
- WERBACH, K. AND HUNTER, D. 2012. *For the Win: How Game Thinking Can Revolutionize Your Business*. Wharton Digital Press.
- WILSON, C., BOE, B., SALA, A., PUTTASWAMY, K. P. N., AND ZHAO, B. Y. 2009. User interactions in social networks and their implications. In *Proceedings of the 4th European conference on Computer systems*. ACM, New York, NY, USA, 205–218.
- XU, Y., CAO, X., SELLEN, A., HERBRICH, R., AND GRAEPEL, T. 2011. Sociable killers: understanding social relationships in an online first-person shooter game. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. CSCW '11. ACM, New York, NY, USA, 197–206.

Online Appendix to: Cheating in Online Games: A Social Network Perspective

JEREMY BLACKBURN, University of South Florida
NICOLAS KOURTELLIS, University of South Florida
JOHN SKVORETZ, University of South Florida
MATEI RIPEANU, University of British Columbia
ADRIANA IAMNITCHI, University of South Florida

A. DATA COLLECTION

A.1. Crawling the Steam Community

Although a fledgling Web API was available in early 2011, it did not expose a method for obtaining the friendslist of users. Thus, using the unmetered, consumable XML on the Steam Community web site, we crawled between March 16th and April 3rd, 2011. The crawler collected user profiles starting from a randomly generated set of SteamIDs and following the friendship relationships declared in user profiles. To seed our crawler, we generated 100,000 random SteamIDs within the key space, of which 6,445 matched configured profiles.

The crawling was executed via a distributed breadth first search. Each of the initial seed SteamIDs was pushed onto an Amazon Simple Queue Service (SQS) queue. Each crawler process popped one SteamID off this queue and retrieved the corresponding profile data via a modified version of the Steam Condenser library [Staudt 2011]. The profile data of the crawled user was stored in a database and any newly discovered users (i.e., friends that were previously unseen) were added to the SQS queue. Crawling proceeded until there were no items remaining in the queue. Using RightScale, Inc's cloud computing management platform to automatically scale the crawl according to the number of items in the SQS queue, we ended up running up to six Amazon "c1.medium" EC2 instances executing up to 15 crawler processes each.

A.2. High Resolution Monitoring of Newly Banned Users

The original version of this work [Blackburn et al. 2012] made use of two crawls, one that took place during March, 2011 and another that took place during October, 2011, as well as a 3rd party database of ban date observations (<http://www.vacbanned.com>) to perform temporal analysis. Unfortunately, the large period of time between our two crawls and the incomplete 3rd party database with unknown collection methodology precluded more sophisticated analysis.

Fortunately, two new Web API methods were made available allowing us to collect a high resolution data. One method provides access to timestamped friends lists and another that provides the ban states (without dates) of up to 100 users at a time. With these API methods we made daily observations of the ban status of users, as well as neighborhood observations of a newly banned users and a control set of non-cheaters.

Ban status observations consist of the id of the observed user, the time stamp the observation was made, and a delta of the ban states. There are three different ban types: 1) **VACBanned**, 2) **CommunityBanned**, and 3) **EconomyBan**. We only make use of the **VACBanned** type in this work (the others are at early stages of being applied and there is no documentation describing their use).

VACBanned is the boolean cheating flag described earlier. **CommunityBanned** is a boolean that we have been unable to find documentation about. However, we discovered that, unlike

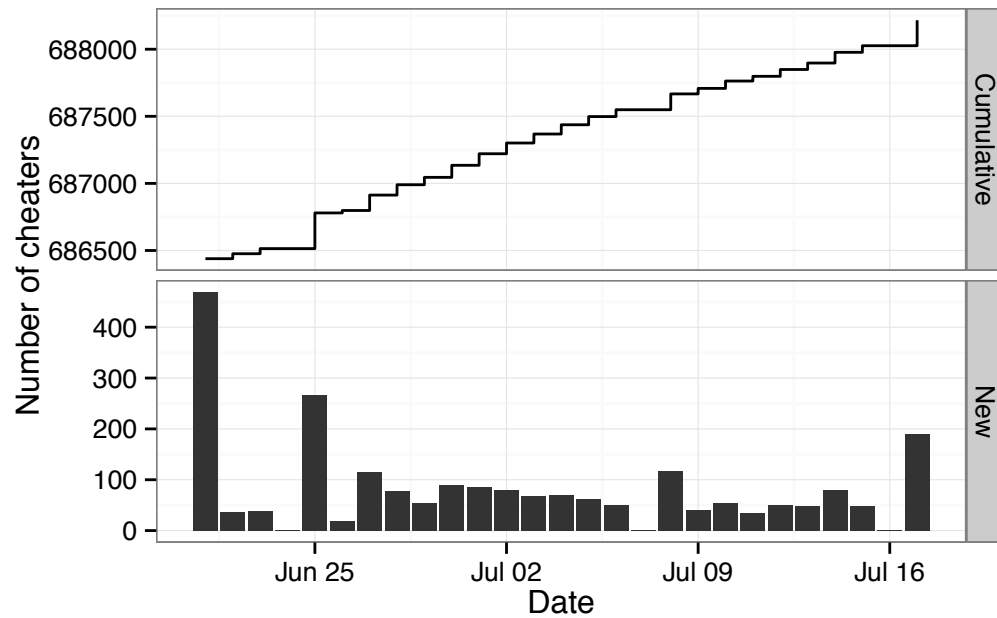


Fig. 13. The number of cheaters in our system, per day from June 21 to July 17, 2012.

a VAC ban, it is not permanent: we observed users having a community ban applied, and later removed. We suspect it bars access to Steam Community features such as commenting on other users' profiles and sending new friend requests. The **EconomyBan** is related to a cross-game³ virtual goods trading platform added to Steam in August 2011 and can take one of three string values: 1) *none*, 2) *probation*, and 3) *banned*. Along with this platform came cheating related to virtual assets [Duh and Chen 2009] and many of the problems discussed in [Keegan et al. 2010], as well as more traditional unscrupulous acts such as bait and switch scams [Ku et al. 2007]. Trade bans are displayed on a user's profile in a similar fashion to VAC bans, but, the three value system allows probated users to be reinstated. To the best of our knowledge, trade bans are applied *manually*, although individual games might very well employ algorithmic techniques to identify, e.g., gold farmers [Ahmad et al. 2011] and bring offenders to the attention of Valve. Although we captured information on all ban types, the remainder of this paper deals only with VAC bans.

On June 20, 2012 we began daily observations of an initial set of 9,025,656 public profiles from the data set described in Section 3.1. In total, we collected over 525 million ban status observations for over 9 million distinct users, with Table II giving the exact numbers. Figure 13 plots the cumulative number of bans per day, as well as the number of new bans in the system from June 21, 2012 to July 17, 2012. We found an average of 83 users were flagged as cheaters per day, but, this number varied quite a bit: from 0 to over 400.

At first we manually ran a set of scripts every day, but on August 7, 2012, the collection was fully automated and we began monitoring the neighborhoods of newly banned users. The system collection system works as follows:

At 12:01 AM EST each morning, all known SteamIDs in the system are randomly ordered and split into groups of 100 IDs. Each group of 100 IDs is entered into a job queue. A set

³Users can trade items from game A for items from game B, and even trade unactivated game licenses

of worker processes pops jobs from this queue and queries the Steam Web API for the ban statuses of each group of IDs. The response from the Web API is pushed onto another queue where worker processes split the aggregated response into a per-user result, which are in turn fed to another queue. An observation delta is then computed for each individual user, and stored in a MongoDB cluster of 10 physical machines (one MongoS, and 3 replica-sets of 3 nodes each, organized into 3 shards). The entire process takes about 6 hours to complete each day.

The multiple queue system is used primarily to help ensure maximum resource utilization, but also emulates a sort of non-blocking I/O for each step of the data collection process. For example, ideally we would collect ban statuses for every user at exactly the same time, however reality dictates that collection is bounded by *at least* the latency the Steam Web API introduces. There is also delay introduced because the web API call works on 100 users at a time, but we must compute deltas a single user at a time. If we were to perform the ban status query and compute the deltas for the returned statuses in sync with the Web API request/response, it would significantly increase the time between each group scan. The queue system, however, decouples the raw data collection and delta computations increasing overall performance as pushing data to the queue is much faster than computing deltas. Of course, the queue system also allows for a multiprocess programming model that is much simpler than threads, and also naturally scales out horizontally to multiple machines.

B. STATISTICAL TESTS USED

We used two tests of statistical significance when comparing distributions in this paper: the two sample Kolmogorov-Smirnov test and the permutation test.

B.1. Two Sample Kolmogorov-Smirnov Test

The KS is nonparametric and uses the supremum (least upper bound) of the set of distances between each point of two empirical distribution functions for its test statistic (D). Unfortunately, while the KS test is distribution agnostic (e.g., the data does not have to be normally distributed) it operates only on *continuous* distributions with no ties. Fortunately, the KS test can be bootstrapped to avoid these issues [Sekhon 2011]. For all reporting in this paper we used 1,000 bootstrap samples.

B.2. The Permutation Test

A permutation test uses random reshuffles of the data to get the distribution of a test statistic under a null hypothesis. In our case the null hypothesis is that of no difference between two groups, cheaters and noncheaters, and the random reassignment of these labels to the elements of the data vector followed by the calculation of the statistic of interest yields a distribution of this statistic, when the reassignment is done many times, against which its observed value can be compared.

The permutation method, also known as a randomization test, is a straight forward and intuitive approach. Consider a 2 column vector representing two distributions where the first column is a label identifying the distribution (e.g., “cheaters” or “noncheaters”) to which the value in the second column belongs. We then calculate a test statistic T^4 .

Operating under the null hypothesis that the labels are meaningless (and thus the distributions are the same), we randomly permute the labels and compute a new test statistic T_i . Due to the random labeling, if we perform the permutation repeatedly, our T_i s should be uniformly distributed, and thus the null hypothesis would have the original T statistic appearing anywhere in the ordered distribution of T_i s with equal probability.

⁴In theory, *any* appropriate test statistic can be used, but we use a standardized linear statistic as described in [Hothorn et al. 2008]

If we perform the permutation M times, we can then calculate a p value as the fraction of permutations where $T_i \geq T$. In other words,

$$p_{permute} = \frac{1}{M} \sum_{i=0}^M I(T_i \geq T)$$

(where I is the indicator function returning 1 if its argument is true, and 0 otherwise). For all reporting in this paper we use $M = 100,000$.