Branded with a Scarlet "C": Cheaters in a Gaming Social Network

Jeremy Blackburn*, Ramanuja Simha∓, Nicolas Kourtellis*, Xiang Zuo*,
Matei Ripeanu‡, John Skvoretz±, Adriana Iamnitchi*
*Department of Computer Science & Engineering, University of South Florida, Tampa, FL, USA
± Department of Sociology, University of South Florida, Tampa, FL, USA
∓ Department of Computer & Information Sciences, University of Delaware, Newark, DE, USA
‡ Department of Elec. & Comp. Engineering, University of British Columbia, Vancouver, BC, Canada
{jhblackb, nkourtel, xiangzuo}@mail.usf.edu, rsimha@cis.udel.edu,
matei@ece.ubc.ca, jskvoretz@usf.edu, anda@cse.usf.edu

ABSTRACT

Online gaming is a multi-billion dollar industry that entertains a large, global population. One unfortunate phenomenon, however, poisons the competition and the fun: cheating. The costs of cheating span from industry-supported expenditures to detect and limit cheating, 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 collected in-game interaction data of over 10 thousand players from a popular multiplayer gaming server. We show that cheaters are well embedded in the social and interaction networks: their network position is largely indistinguishable from that of fair players. We observe that the cheating behavior appears to spread through a social mechanism: 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. Also, we observe that there is a social penalty involved with being labeled as a cheater: cheaters are likely to switch to more restrictive privacy settings once they are tagged and they lose more friends than fair players. Finally, we observe that the number of cheaters is not correlated with the geographical, real-world population density, or with the local popularity of the Steam Community.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*Information networks*; K.4.2 [**Computers and Society**]: Social Issues—*abuse and crime involving computers*

Keywords

cheating in online games, social network analysis

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 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 [3].

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 [18]". 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). 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 [19, 20]. For other cheaters, a competitive advantage and the desire to win is motivation enough [21].

Cheating is seen by the game development and distribution industry as both a monetary and a public relations problem [9] 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 non-secured servers, 26 were servers owned and administrated by a competitive gaming league that operates its own anti-cheat system.

Gaming interactions mimic, to some extent, real-world interactions [27]. Understanding cheaters' position in the social network that connects gamers is relevant not only for evaluating and reasoning about anti-cheat actions and policies in gaming environments, but also for studying social networks at large. Studying cheaters can serve to better understand the behavior of individuals that abuse the shared social space in large-scale non-hierarchical communi-

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.
WWW 2012, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1229-5/12/04.

ties. 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.

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 users. 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, and highlights the social consequences of the publicly visible cheating flag.

Our study shows that cheaters are well embedded in the social network; they exhibit a high degree of homophily; their geo-social characteristics differ from those of fair players; and while the cheating flag does not affect their aggregate well being in the gaming environment, it is penalized by friendship loss, and marked by some degree of embarrassment. Additionally, our temporal analysis of the cheating data suggests that cheating behavior spreads via a social mechanism: 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.

An overview of related work is presented in Section 2. Our datasets are presented in Section 3, along with our data collection methodology. Section 4 analyzes the position of cheaters in the network from the perspective of declared relationships, in-game interactions, and the strength of their relationships measured via social-geographical metrics. It also presents the effect of the VAC-ban on individual players. Section 5 reasons about possible mechanisms for spreading the cheating behavior. Section 6 concludes with a summary of our findings and their consequences.

2. RELATED WORK

Cheating in social gaming is a relatively unexplored area. Nazir et al. study fake profiles created to gain an advantage in social gaming contexts in [21]. Through 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 creating fake profiles to perform a Sybil attack, whereas cheaters in Steam Community are not trying to alter the structure of the social graph. Instead, they attack 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. [19] 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. [1] further examined trade networks of gold farmers and propose models for deviant behavior prediction.

Their data set 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 Steam Community players. 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 [21], and multiple orders of magnitude more cheaters than players in [19, 1].

Although not much quantitative analysis has been performed, cheating in video games has been studied qualitatively. Duh and Chen describe several frameworks for analyzing cheating, as well as how different cheats can impact online communities in [11]. Dumitrica examines *Neopets*, a web based social game in [12]. 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.

Social networks of online gamers have been addressed in recent studies. Szell and Thurner [27] provide a detailed analysis of the social interactions between players in Pardus, a web-based Massively Multiplayer Online Game. They employ, as we do in this study, traditional tools from social network analysis, however, there are significant differences in the datasets used. First, declared relationships between users in the Steam Community are *informed* by underlying in-game interactions, but exist in a more general "gaming" context than in Pardus, where the social network is built on interactions within the context of one game. Second, players in Pardus can declare friends and enemies. In our study, players can only declare friends. While we do provide some results based on interaction data from a Team Fortress 2 (TF2) server, TF2 players do not declare friends and foes, but compete on ad-hoc, opposing teams. Finally, while there might be cheaters present in the Pardus dataset, they are not identified or studied.

Xu et al. [29] interviewed 14 Halo 3 players to study the meaning of relationships within an online gaming context. 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. In addition to in-game interactions on a single game server, we also measure and analyze the social structure of millions of online gamers and their relationships with the deviant class of users that cheat.

General gaming studies have characterized network traffic due to gaming, resource provisioning, work load prediction, and player churn in online games [7, 8, 13, 14, 15]. Other studies have focused on the psychological and social properties of gamers [16] and gaming communities [6, 28, 10].

3. DATASETS

Steam controls between 50% and 70% of the PC digital download market [25], and claims over 30 million user accounts as of October 2011. Steam is run by Valve, who also develops some of the most successful multiplayer first-person shooter (FPS) games.

While games from a number of developers and publishers are available for purchase on Steam, an important 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 multiple individually-owned and operated servers. Because there is no central entity controlling game play and a very large number of servers to choose from, the communities that form around individual servers are essential to the prolonged health of a particular game.

3.1 The Steam Community

Recognizing the social nature of gaming in general, Valve created 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 client 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 stymie the creators and distributors of cheats and hacks, the details of how VAC works are not made public. What is known is that 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.

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 chatting with other players.

3.2 Data Collection

In our analysis we used three data sources. The vast majority of our data was obtained by crawling the Steam Community website to collect user profiles and the resulting social network. In order to augment profile information with the (approximate) time of VAC bans, we queried the `vacbanned.com` site. And finally, we obtained in-game interactions from a TF2 server located in California.

Crawling the Steam Community: Using unmetered, consumable XML on the Steam Community web site, we crawled during 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 (64-bit identifiers with a common prefix that reduced the ID space to less than 10^9 possible IDs), 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. 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 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, gameplay statistics for the past two weeks, a user-selected geographical location, and a flag (VAC-ban) that indicates whether the corresponding user has been algorithmically found cheating. We augmented the information for the VAC-banned players with a timestamp that signifies when the VAC ban was first observed (as explained next).

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 1 gives the exact numbers.

Collecting VAC Ban Timestamps: We collected historical data on when a cheating flag was first observed from a 3rd party service, `vacbanned.com`, that allows users to enter a SteamID into a search box to check whether or not that SteamID has been banned. If the account is banned, the date the ban was first observed is provided. We also re-crawled (between October 18th and October 29th 2011) all Steam profiles discovered during the first crawl without a VAC ban, to identify which non-cheaters had been flagged as cheaters since April 2011. Of these, $43,465$ now have a VAC ban on record.

`Vacbanned.com` had observed ban dates for 423,592 of the cheaters we discovered during our initial crawl. Figure 1 shows a CDF of these ban observations over time. The earliest dates indicate users that were banned prior to December 29th, 2009. We combined the "banned-since" dates from our original crawl, `vacbanned.com`, and our re-crawl. In the case of a user profile having more than one ban date (due to the 3 sources), the earliest date was chosen. It is important to note that all ban dates were treated as "on or before" as opposed to a precise timestamp. This is because the ban dates are when the ban was first observed by a 3rd party (`vacbanned.com` or our crawler), not necessarily when it was applied by Valve.

In-game interactions: We have acquired detailed game play logs of a 32-simultaneous player VAC-secured TF2 server

Account	*All*	*Edges*	*Profiles*	*Public*	*Private*	*Friends only*	*With location*
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

Table 1: Size of the *Steam Community* dataset.

located in California. Our logs span just over 2 months from April 1 to June 8, 2011, and consist of various game-specific events involving 10,354 players. Because this server is VAC-secured, no players that have cheated in TF2 appear in the logs; the only cheaters that appear are those that were caught in a different game.

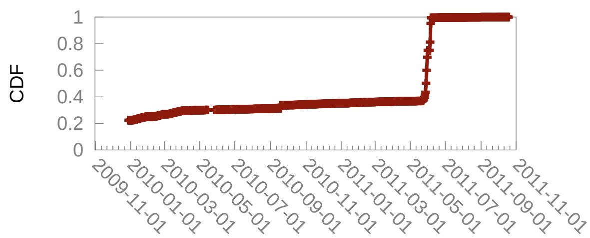


Figure 1: Historical VAC ban dates as reported by `vacbanned.com`. The date of discovery is on the x-axis, and the cumulative probability distribution of the number of discovered accounts is on the y-axis. The jump around end of May 2011 is probably due to an effort from the website to populate their database.

From the server logs we extracted 5 different interactions types between users, and constructed an interaction graph where an edge exists between two players if they interacted together during the game. The resulting graph contained 10,354 users (93 were cheaters) and had 486,808 edges.

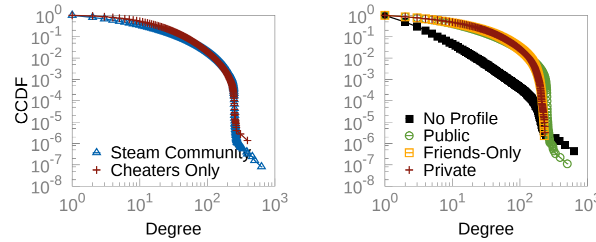
4. CHEATERS AND THEIR FRIENDS

One line of thought in moral philosophy is that (non)ethical behavior of an individual is heavily influenced by his social ties [23]. Under this theory, cheaters should appear tightly connected to other cheaters in the social network. On the other hand, unlike in crime gangs [1], cheaters do not need to cooperate with each other to become more effective. Moreover, playing against other cheaters may not be particularly attractive. 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 four axes: First, we explore the relationship between cheating status and the number of friends a user (Section 4.1); second, we use the in-game interaction trace to understand who plays with cheaters (Section 4.2); third, we try to understand whether cheaters are visibly penalized by the other members of the social network (Section 4.3); and, finally, we explore the relationship between social network proximity and two other proximity metrics, geographical and community-based (Section 4.4).

4.1 Who is Friends with Cheaters?

The degree distribution of the Steam Community graph as a whole, just cheater profiles, as well as for private, friends-only profiles, and users without profiles are plotted as CCDF in Figure 2. For users without a profile or private profiles,



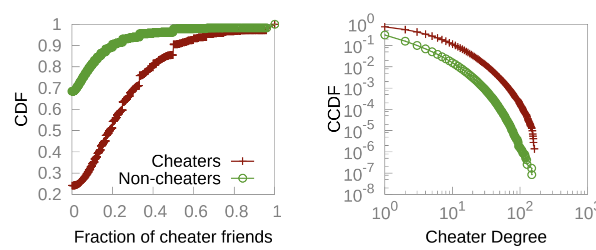
(a) All users and cheaters. (b) Public, private, friends only, and no profile.

Figure 2: Degree distributions in Steam Community

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. 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. We thus make two observations. First, cheaters have about the same number of declared friends as non-cheaters. Second, this data highlights that attempting to hide connection information through *private* or *friends-only* profile privacy settings is unsuccessful: the player's position in the social network is revealed by the privacy settings of his friends.



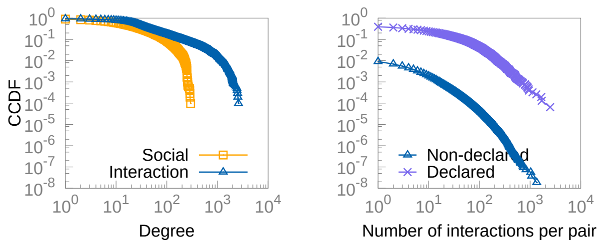
(a) (b)

Figure 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 cheaters' friends that are cheaters vs. the number of non-cheaters' friends that are cheaters.

While cheaters are mostly indistinguishable from fair players using the node degree distribution, a more important question is whether their deviant 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 cheaters friends for both cheaters and non-cheaters. This figure is comparable to Figure 2(a), but displays only the "cheating" degree of users.

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 is statistically significant by using the Kolmogorov-Smirnov (KS) test to verify that the two samples are drawn from different probability distributions ($p < 0.01$, $D = 0.4367$, and $p < 0.01$, $D = 0.3787$ at the 5% significance level for Figures 3(a) and 3(b), respectively).

4.2 Who Plays with Cheaters?



(a) (b)

Figure 4: (a) CCDF of the declared friendship degree and interaction degree of users from a popular TF2 server. (b) CCDF of the number of interactions between declared and non-declared pairs of friend players on a popular TF2 server.

To investigate if the declared friendships reflect in-game interactions and if cheaters have similar playing patterns with non-cheaters, we studied the 2-month interaction network generated from the TF2 server logs. Figure 4(a) plots a CCDF of the declared friendship degree as well as the interaction degree for players appearing in the interaction network. We first note that even on a single server for a single game, players generally interact with considerably more players than they have declared friendships with. However, the correlation between the number of declared friends and the number of distinct interaction partners is low (Pearson coefficient 0.16). This suggests that being popular in the social network does not necessarily translate to an increase in unique interaction partners.

Figure 4(b) compares the number of interactions between declared friends and players who are not declared friends. The plot suggests that players with a declared friendship interact with each other more often than players without a declared friendship, indicating that Steam Community friendships are representative of in-game interactions.

But, are cheaters ostracized by non-cheaters during gameplay? We answer this question by examining how cheaters interact with declared and non-declared friends. Figure 5 plots the CCDF of the number of pairwise interactions, with at least one member of the pair being a cheater (or non-cheater), for both declared and non-declared Steam Community friendships.

There are two important observations that result from these plots. First, we see that cheaters, like the population

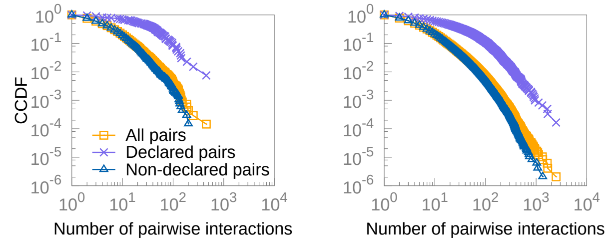


Figure 5: CCDF of the number of interactions between declared and non-declared pairs of players, with at least one user in the pair being a cheater (left) or a non-cheater (right), on a popular TF2 server.

of the server as a whole, are likely to have more interactions with declared friends. Second, we notice that interacting pairs with at least one cheater in the pair have fewer absolute interactions than the server as whole.

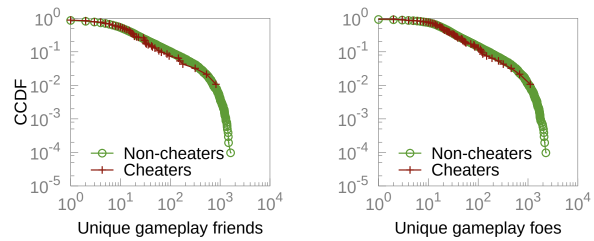


Figure 6: CCDF of the number unique gameplay friends and foes of cheaters and non-cheaters.

Since TF2 games are between competing teams, players on the same team can have cooperative interactions, and those on opposing teams can have antagonistic interactions. If players had overwhelmingly negative feelings towards cheaters, one might expect cheaters to be involved in fewer cooperative interactions than antagonistic interactions, i.e., players might target those with a cheating flag in a form of vigilante justice. This does not seem to be the case as demonstrated by Figure 6, which presents the CCDF of number of unique "friend" (cooperative) and "foe" (antagonistic) partners for cheaters and non-cheaters. While cheaters tend to have slightly fewer unique gameplay partners than non-cheaters, the difference is negligible (the KS test was unable to reject the null hypothesis with $p = 0.7498$, and $p = 0.7602$ respectively), indicating that the cheater/non-cheater status of a player does not hold much weight during active game play. There are several explanations for this finding. First, it is possible that the perception of a general distaste for cheaters is only the effect of a vocal minority. Second, the lack of a visible *in-game* cheater brand allows cheaters to blend in with the rest of the players. Finally, gamers might be tolerant of cheaters as long as they are not actively displaying cheating behavior.

4.3 Are Cheaters in Disgrace?

While aggregate-level information shows little differentiation between cheaters and non-cheaters, the effect of the VAC-ban mark can better be understood by analyzing the transition from non-cheater to cheater. We answer the following two questions: 1) Are cheaters shamed by the mark on their permanent record? and 2) Does the community shun cheaters once their transgressions are revealed?

Of the new cheaters discovered from our re-crawl, 87% had no change in privacy state, and nearly 10% changed their privacy setting from public to a more restrictive setting. In comparison, in our control group of re-crawled non-cheaters, privacy settings remained unchanged for over 97% of users, and less than 3% changed to a more restrictive setting. Cheaters seem to value their privacy more once their sins are laid bare, perhaps in the naive hope that a more restrictive setting will provide a measure of protection from a potentially disapproving community.

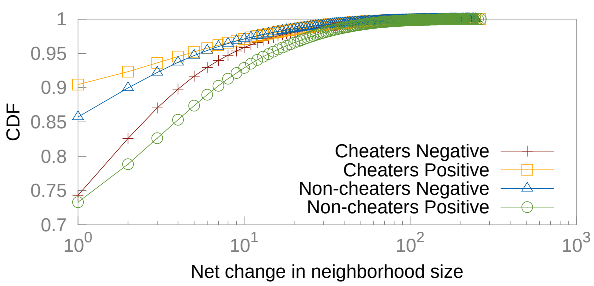


Figure 7: CDF of net change in cheaters' and non-cheaters' neighborhood size.

But is the community disapproving? Figure 7 plots the CDF of net change in the number of friends for cheaters and non-cheaters during the six months between our two crawls. Of the still public cheaters in our re-crawl, 44% had a net loss of friends, 13% had a net gain, and 43% had no change. Of the non-cheaters in our new crawl, 25% had a net loss of friends, 36% had a net gain, and 39% had no change. While both sets of users exhibited fluctuations in the size of their neighborhoods, more cheaters lost friends and fewer cheaters gained friends when compared to non-cheaters. Treated as a whole, cheaters lost nearly twice as many friends as they gained, and non-cheaters gained twice as many friends as they lost. Overall, non-cheaters continue to gain friends, and cheaters, while not overtly ostracized, appear to have trouble making new friends and may lose a few of their previous ones.

There are several possible explanations for the changes in neighborhood sizes we observe. First, evidence suggests that online gamers "clean up" their friend lists to make room for new friends, removing people they no longer play with [29]. However, because so few users are near the 250 friend limit (as seen in Figure 2(a)), we do not believe this is the primary contributing factor to neighborhood size fluctuations. A second explanation is that the Steam client, by default, issues "pop up" notifications that are visible in game whenever a friend starts playing any game. If a gamer has many active friends, these pop ups might become distracting, possibly prompting a purge of friends they no long actively play with regardless of how close they are to the friend limit. A final explanation, especially with respect to the net loss of friends cheaters suffer, is that cheaters are deliberately severing their ties once they are caught cheating. We observed one account in particular that went from 200 to 0 friends after the VAC ban was issued. "Social suicide" might account for large decreases in degree, as it is far more probable that the cheater himself deletes friends, rather than each of his friends deleting the cheater.

4.4 Are Cheaters Close?

We use two metrics, one geographical, and another social, to quantify the strength of the relationship between Steam Community users.

Geographical proximity: Exploring the relationship between geographical and social-network proximity may give quantitative support to the theory proposed in [12] according to which opinions on cheating are culturally derived. Although framed mainly in the context of a specific set of game rules, we extend the theory into the real world by first observing that user population on Steam Community does not follow real-world geographic population and, more importantly, cheaters are not uniformly distributed. Figure 8 shows *Steam Community* 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. 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.

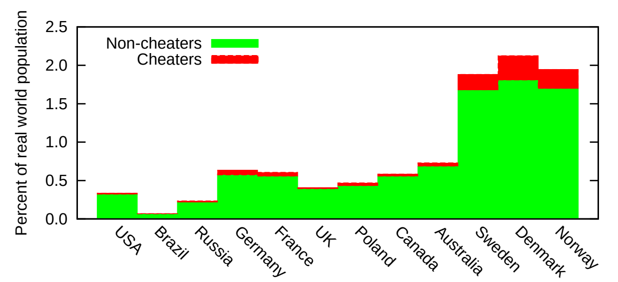


Figure 8: User and cheater populations per country normalized to real world population of that country. The countries are sorted on the x-axis in decreasing order of their real-world populations.

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 [24]. 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. 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 2. We note that a subgraph composed solely of cheater-to-cheater relationships (C-C) has a lower mean distance between nodes and average link length than the location network as a whole.

Figure 9 plots the CDF of node locality for the location network (Steam Community), the cheater-to-cheater sub-

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

Table 2: 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 $\langle NL \rangle$. The FourSquare and BrightKite properties are from [24].

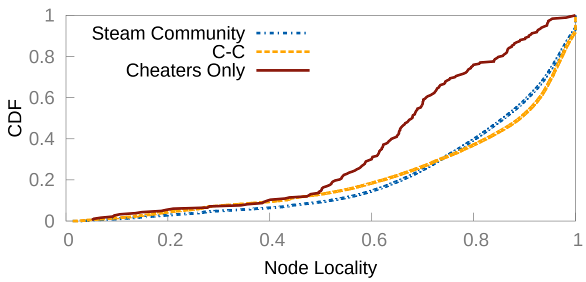


Figure 9: CDF of node locality.

graph (C-C), as well as just the cheaters within the location network (Cheaters Only). 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 [24]. 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.

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 [22] 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.

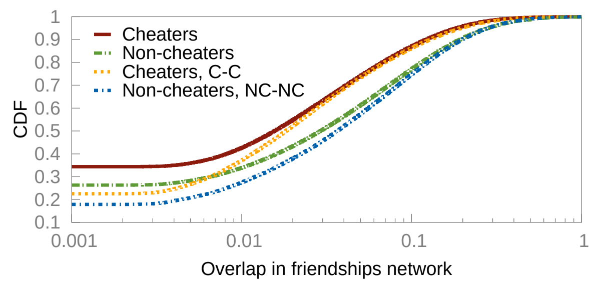


Figure 10: CDF of social proximity for cheater and non-cheater pairs when we consider all relationships, only cheater to cheater relationships (labeled C-C) and only non-cheater to non-cheater relationships (labeled NC-NC).

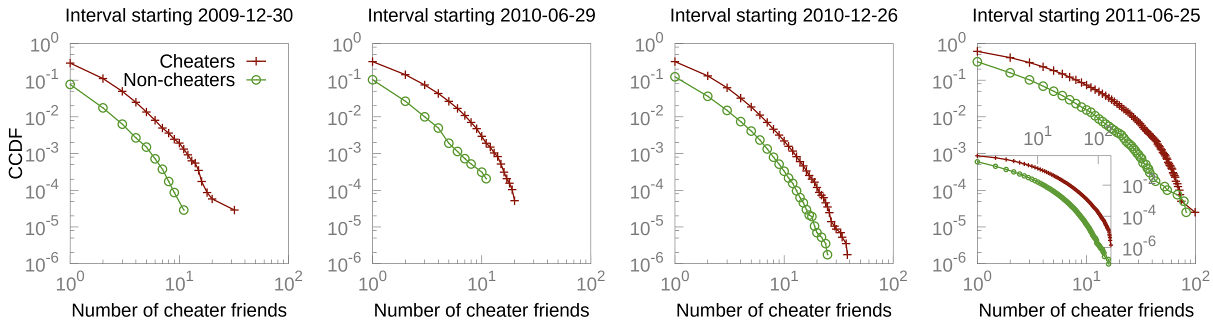
Figure 10 shows a higher overlap for cheater pairs in the cheater-only graph and non-cheater pairs in the non-cheater-only graph compared to the respective overlaps in the overall social network. This suggests that social relationships are weaker between different types of players (cheaters to non-cheaters) than within a uniform group.

5. PROPAGATION OF CHEATING

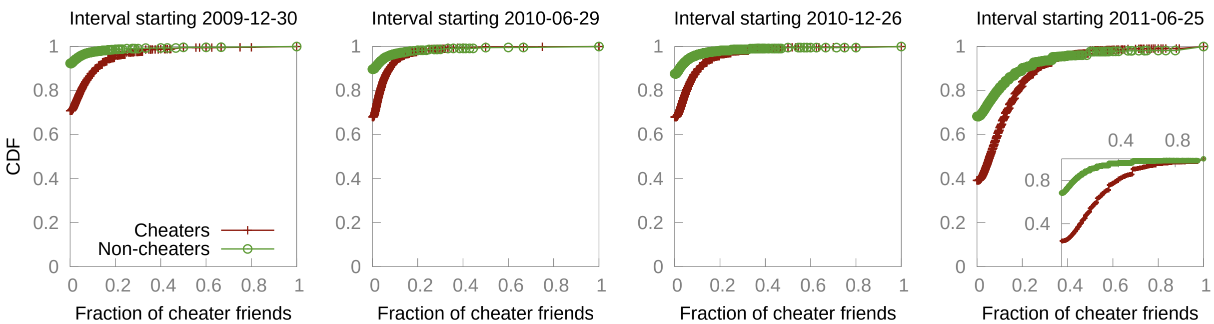
How does cheating behavior spread in the Steam Community? A first insight can be obtained by investigating how cheating bans propagate in the network over time (Section 5.1). Further insight can be obtained by understanding whether cheaters hold positions of influence in the social network (Section 5.2).

5.1 How Does Cheating Propagate over Time?

Based on the observations from Figure 3, we hypothesize that the friends of known cheaters are at risk of becoming cheaters themselves. To test this hypothesis, we explored whether cheaters discovered during a given time interval were more likely to be friends of previously discovered cheaters than to be friends with non-cheaters. Again, we stress that banned dates must be treated as "on or before"



(a) CCDF of the number of cheater friends of newly discovered cheaters and a random sample of non-cheaters.



(b) CDF of the fraction of cheater friends of newly discovered cheaters and a random sample of non-cheaters.

Figure 11: The spreading of cheating behavior in the Steam Community over four 180-day time intervals. The inset plot shows the final state of the network. A randomly selected control sample of non-cheaters is used for comparison in each time interval.

as opposed to exact timestamps. To mitigate the effects of this uncertainty, we chose to examine cheaters discovered over 180-day long intervals.

We begin by assuming that all users in the Steam Community friendship network are non-cheaters. We then initialize the network by marking the 94,522 users found to have a VAC ban on or before December 29, 2009 (i.e., the earliest date retrieved from vacbanned.com). For the first 180-day interval (between December 30, 2009 and June 28, 2010), 34,681 players were found to have a VAC ban. For these users, we calculated and plotted the number and fraction of their cheater friends (i.e., from the 94,522 cheaters found previously). We repeat these steps for another 3 time intervals, with 19,294, 571,975, and 43,465 cheaters found in each. The third interval (starting at December 26, 2010) contains the bulk of cheaters, since their VAC ban was first observed by our initial crawl (and not from `vacbanned.com`). However, as shown next, the differentiation between cheaters and non-cheaters holds true for all intervals. In addition to a best effort approximation of the timestamp of the VAC bans, the data constructed this way has another caveat: the social network is from our March/April 2011 crawl, but we show in Section 4.3 that the network is quite dynamic. We verified that despite the change in number of friends over time, the trend is preserved: we recalculated the fraction of cheaters in the 1 hop neighborhoods of users based on the state of their relationships as determined by our October 2011 re-crawl and we found the non-cheaters CDF to dominate the cheaters CDF as in Figure 3(a).

Figure 11 plots the results of our experiments. Each subplot represents only the cheaters discovered during the corresponding time interval, and an equal number of randomly sampled non-cheaters (statistical significance confirmed by KS test at 5% significance level, $p < 0.01$ in all cases, and D = 0.215, 0.2172, 0.1963, and 0.2911 for the fraction of cheater friends distributions of each interval, respectively). From the plots we see evidence that users with both a higher absolute number of cheater friends, as well as those with proportionally more cheater friends are more likely to become cheaters themselves.

To understand if the number of cheater friends has any predictive power on the state of a player, we used the framework developed by Backstrom et al. [5]. We first created a dataset of fair players with at least one cheater friend at the time of our first crawl. We then labeled the players from this set who were marked as cheaters by the time of our second crawl. We consider the new cheaters as joining a group in the sense of the Backstrom study to estimate the probability of a fair player becoming a cheater. We then applied the decision tree technique, using the number of cheater friends as the single feature, and achieved a ROCA of .61. With perfect knowledge, that is, if every single cheater was marked in our dataset by a perfect and timely cheating detection system, we would expect this value to be higher. (In comparison, Backstrom et al. show a ROCA value of .6456 for the DBLP communities and .69244 for LiveJournal.)

While these results show that the number of cheater friends has predictive power for the transition of a player from fair to cheater, it is difficult to ascertain the specifics of how the behavior propagates. Specifically, it is difficult to distinguish between homophily and contagion by simply observing the network properties and without strong assumptions about the social process [4, 26]. However, an individual's propensity towards unethical behavior, and cheating in particular, has been shown to be dependent on social norms, including the saliency of the behavior, and whether or not the behavior is observed from in- or out-group members [17]. Our large-scale findings further corroborate this theory and suggest that cheating behavior is not preponderantly dependent on a personal strategy that takes into account player-local information only, but spreads through a contagion process.

5.2 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 relative importance of 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 [2].

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 3 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 4.3 could provide an explanation for this. There seems 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.

6. SUMMARY AND DISCUSSION

Online gaming has recently 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.

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

Table 3: Percentage of cheaters found in top-N% of high degree centrality (DC) and betweenness centrality (BC) users in the Steam Community.

Due to the scale of our dataset, the majority of our computations used the MapReduce framework via the `python mrjob` interface for Hadoop on Amazon Elastic MapReduce. Our MapReduce stages involved graph pre-processing, gameplay statistics computations, geographical data processing, computing degree distributions, intersections of sets, and geo-social metrics. Each solution included several MapReduce pipelines (chains of map tasks and reduce tasks) of smaller subtasks.

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 both the social network and interaction logs from a multiplayer gaming server, where cheaters were not targeted or treated overly different from non-cheaters.

However, when we examine the transition from fair player to cheater, we observe the effects of the cheating brand. First, cheating behavior appears to spread through a social 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. 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. Finally, we 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 legislature and in the 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.

Acknowledgements

We thank Y. Boshmaf, F. Figueiredo, A. Sotirkopoulos, S. Al-Kiswany, and C. Shawn of University of British Columbia for their feedback on an earlier version of this work. We also thank the anonymous reviewers for their input. This research was supported by NSF under Grants No. CNS 0952420 and CNS 0831785.

7. REFERENCES

[1] M. A. Ahmad, B. Keegan, S. Sullivan, D. Williams, J. Srivastava, and N. Contractor. Illicit Bits: Detecting and Analyzing Contraband Networks in Massively Multiplayer Online Games. *SocialCom'11*, 2011.

[2] T. Alahakoon, R. Tripathi, N. Kourtellis, R. Simha, and A. Iamnitchi. K-path centrality: A new centrality measure in social networks. In *SNS 11*, 2011.

[3] APB Reloaded Dev Blog. `http://goo.gl/m9klr`, 2011.

[4] S. Aral, L. Muchnik, and Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Science*, 2009.

[5] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006.

[6] M. Balint, V. Posea, A. Dimitriu, and A. Iosup. An Analysis of Social Gaming Networks in Online and Face to Face Bridge Communities. In *LSAP 11*, 2011.

[7] C. Chambers, W. Feng, S. Sahu, D. Saha, and D. Brandt. Characterizing online games. *IEEE/ACM TON*, 2010.

[8] M. Claypool, D. LaPoint, and J. Winslow. Network analysis of counter-strike and starcraft. In *IPCCC 03*, 2003.

[9] M. Consalvo. *Cheating: Gaining Advantage in Videogames*. MIT Press, Cambridge, Massachusetts, 1st, edition, 2007.

[10] N. Ducheneaut, N. Yee, E. Nickell, and R. Moore. The life and death of online gaming communities: a look at guilds in world of warcraft. In *CHI 07*, pages 839–848, 2007.

[11] H. B.-L. Duh and V. H. H. Chen. Cheating Behaviors in Online Gaming. *Online Communities and Social Computing*, 2009.

[12] D. D. Dumitrica. An exploration of cheating in a virtual gaming world. *Journal of Gaming & Virtual Worlds*, 2011.

[13] W. Feng, F. Chang, W. Feng, and J. Walpole. Provisioning on-line games: a traffic analysis of a busy counter-strike server. In *IMW 02*, pages 151–156, 2002.

[14] W. Feng, F. Chang, W. Feng, and J. Walpole. A traffic characterization of popular on-line games. *IEEE/ACM TON*, 13:488–500, 2005.

[15] W. Feng and W. Feng. On the Geographic Distribution of On-line Game Servers and Players. In *NetGames 03*, 2003.

[16] T. Fritsch, B. Voigt, and J. Schiller. Distribution of online hardcore player behavior: (how hardcore are you?). In *NetGames 06*, 2006.

[17] F. Gino, S. Ayal, and D. Ariely. Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel. *Psychological Science*, 2009.

[18] J. Huizinga. *Home Ludens: A Study of the Play-Element in Culture*. Beacon Press, Boston, Massachusetts, reprint edition, 1950.

[19] B. Keegan, M. Ahmed, D. Williams, J. Srivastava, and N. Contractor. Dark Gold: Statistical Properties of Clandestine Networks in Massively Multiplayer Online Games. In *SocialCom 10*, pages 201–208, 2010.

[20] Y. Ku, Y. Chen, K. Wu, and C. Chiu. An empirical analysis of online gaming crime characteristics from 2002 to 2004. In *PAISI 07*, pages 34–45, 2007.

[21] A. Nazir, S. Raza, C. Chuah, and B. Schipper. Ghostbusting facebook: detecting and characterizing phantom profiles in online social gaming applications. In *WOSN 10*, 2010.

[22] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.

[23] D. Parfit. *Reasons and Persons*. Oxford University Press, 1984.

[24] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora. Distance matters: geo-social metrics for online social networks. In *WOSN 10*, 2010.

[25] T. Senior. Steam controls up to 70% of the pc download market and is "tremendously profitable". `http://goo.gl/MjgA3`, 2011.

[26] C. R. Shalizi and A. C. Thomas. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research*, 2011.

[27] M. Szell and S. Thurner. Measuring social dynamics in a massive multiplayer online game. *Social Networks*, 32:313–329, 2010.

[28] X. Wei, J. Yang, L. Adamic, R. Araújo, and M. Rekhi. Diffusion dynamics of games on online social networks. In *WOSN 10*, 2010.

[29] Y. Xu, X. Cao, A. Sellen, R. Herbrich, and T. Graepel. Sociable killers: understanding social relationships in an online first-person shooter game. In *CSCW 11*, pages 197–206, 2011.