

Cheaters in a Gaming Social Network

Jeremy Blackburn¹, Ramanuja Simha¹, Clayton Long¹, Xiang Zuo¹, Nicolas Kourtellis¹,
John Skvoretz², Adriana Iamnitchi¹

¹Department of Computer Science and Engineering, ²Department of Sociology
University of South Florida, Tampa, FL, USA

{jblackb, rsimha, cslong2, xiangzuo, nkourtell}@mail.usf.edu, jskvoretz@usf.edu, anda@cse.usf.edu

1. INTRODUCTION

The popularity of online gaming has raised significant interest in understanding the technological needs for supporting gaming platforms. Consequently, various studies characterized network traffic due to gaming, resource provisioning, work load prediction, and player churn in online games [1]. Other studies have focused on the psychological and social properties of gamers and gaming communities [5]. Gamers form strong communities around their activity, with recent work suggesting that explicitly defined relationships of gamers tend to be supported by real world relationships [6].

One major problem with gaming is cheating. 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 [2]. For other cheaters, a competitive advantage and the desire to win is motivation enough [3].

In this work, we analyze how cheaters are embedded in the *Steam Community*, a large online social network for gaming with millions of active users. Cheaters are identified by an automated mechanism operated by the Steam gaming service and their profiles are permanently flagged in a publicly visible way. We analyze differences in gaming-specific properties of cheaters and non-cheaters. We also examine whether cheaters occupy a particular position in the social network when compared to non-cheaters, and characterize the relationships they have among themselves, and with non-cheaters.

Our study shows that cheaters are well embedded in the social network, having similar connectivity to non-cheaters; their social position is largely undistinguishable from that of fair players; and their geographic distribution does not correspond to real world population nor is it determined by the popularity of the Steam Community at a given real world location.

2. DATA COLLECTION

At the time of the data collection, a rate limited web API for accessing the Steam Community was available. Unfortunately, in addition to being limited to 100,000 calls per day, the web API does not provide access to the list of a user's friends, but rather only to limited summary information for a user. As an alternative, Steam Community data are made available via unmetered XML. We built a crawler that performed a distributed breadth first search using up to 5 'c1.medium' Amazon EC2 instances. The crawler's architecture is shown in Figure 1. Using the XML interface,

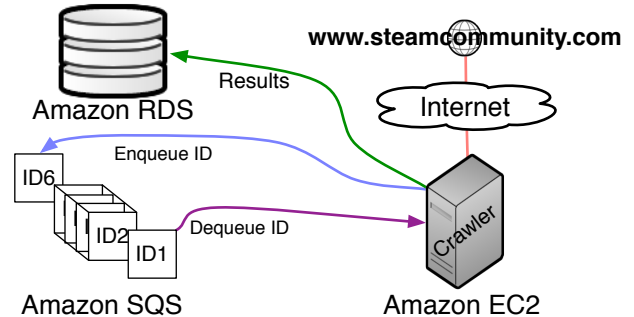


Figure 1: Crawler architecture. Up to 5 ‘c1.medium’ Amazon EC2 instances perform a distributed breadth first search. An Amazon SQS Queue manages discovered, but unvisited, nodes. Crawler output is stored in a database.

we began crawling on March 16th, 2011 and finished crawling on April 3rd, 2011. The majority of the data (over 75%) was collected between March 25th and March 31st.

2.1 Dataset

A user profile consists of a *SteamID*, set of friends (identified by *SteamIDs*), games owned, gameplay statistics for the past two weeks, a user-selected geographic location, and a flag that indicates whether the corresponding user has been found cheating.

Game ownership: Each game has a unique identifier. Games can be added to a user's account in several ways: 1) via the Steam eCommerce and distribution platform, 2) via a registration key provided by certain games purchased from other sources, and 3) via gifting games purchased through Steam to other users.

Gameplay statistics: One of the features of Steam Community is the statistical tracking of gameplay history. The number of hours played in the past two weeks across all games on account is made available. This aggregate statistic is also broken down to a per game statistic, thus showing the number of hours each individual game was played.

Cheating flag: The specifics of what constitutes cheating vary over games, but in general, cheating is behavior that is well outside any reasonable interpretation of the rules of a given game, such as modifying game code to allow a player to see through walls, or to automatically aim for a player.

From an initial 6,445 seeds for *SteamIDs*, we discovered just about 12.5 million users, of which 10.2 million users had a profile configured (about 9 million public, 313 thousand

User Class	N	K	$N_{profile}$	N_{pub}	N_{priv}	N_{fo}	N_{loc}	N_{activ}
All users	12,479,765	88,557,725	10,191,296	9,025,656	313,710	851,930	4,681,829	3,290,325
Cheaters	-	-	720,469	628,025	46,270	46,174	312,354	82,437

Table 1: Properties of the Steam Community dataset: the number of users N , the number of declared friendships K , the number of users with a profile configured $N_{profile}$, the number of public profiles N_{pub} , the number of private profiles N_{priv} , the number of friends only profiles N_{fo} , the number of users with a location specified N_{loc} , the number of users that have played at least 1 game in the past two weeks N_{activ} .

private, and 852 thousand visible to friends only). There are 88.5 million undirected *friendship* edges. Of the users with public profiles, 4.7 million had a location specified, 3.2 million have played at least one game in the two weeks prior to our crawl. 720 thousand users are flagged as cheaters. Table 1 summarizes the dataset properties. From the dataset we constructed a graph consisting of nodes representing users and edges representing the declared friendships between them.

3. CHEATERS IN THE COMMUNITY

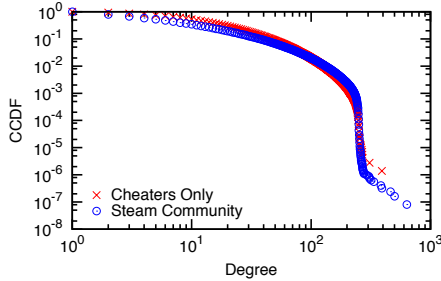


Figure 2: Degree distributions.

The degree distribution of users in Steam Community as a whole, and just the users that are cheaters, are plotted as a CCDF in Figure 2. From the degree distributions we discovered that a hard limit of 250 friends is in place. 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 is also flagged as a cheater. The degree distributions show that cheaters are not pariahs; they exhibit the same social connectivity as Steam Community in whole.

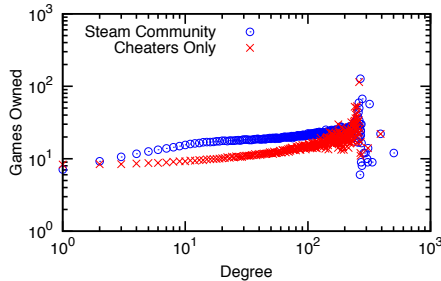


Figure 3: The number of games owned as a function of degree.

Figure 3 shows the average number of games owned as a function of degree on a log-log scale. We observe that there

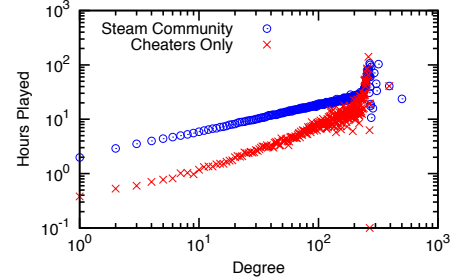


Figure 4: The number of hours played in the past two weeks as function of degree.

is a quick rise in the number of games owned up to about 10 friends, after which we continue to see an increasing relationship at a slower rate. With cheaters, the initial increase in games owned is not as apparent, but the generally increasing relationship still exists. The number of hours played as a function of degree, as seen in Figure 4, exhibits a similar pattern, with a more acute response from the cheaters.

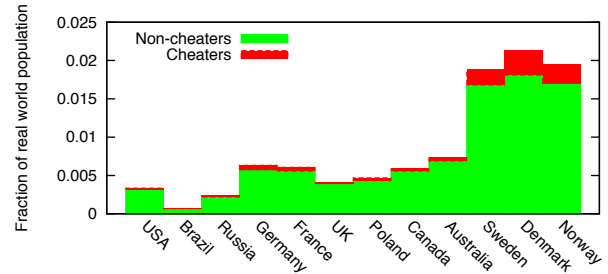


Figure 5: User and cheater populations per country normalized to real world population. The countries are listed on the x-axis by decreasing real world population.

Figure 5 shows Steam Community populations for the 12 countries (5% of all user specified countries) comprising the union of the top ten user populations and the top ten cheater populations. The countries are arranged along the x-axis in decreasing order of their real world populations. The number of Steam Community non-cheaters and the number of cheaters per country are normalized to the real world population of the country. We note that neither Steam Community users as a whole, nor cheaters specifically, are uniformly distributed with respect to real world populations. For example, we found enough profiles to account for nearly 2.1% of Denmark’s 5.5M residents, and cheaters account for nearly 0.32% of the real world population. Accounts located in the USA comprise less than 0.34%, and cheaters less than 0.014% of the 300M real world population. In total, 33% of cheaters are located within these 12 countries.

While Figure 5 shows cheaters are clustered in a few geographically distinct regions, it does not say anything about the effects that distance has on *relationships*. To address this, we measure node locality and geographic clustering coefficients from [4]. Node locality measures how close an individual is to all of her friends and geographic clustering coefficient measures the geographic closeness of triangles of users.

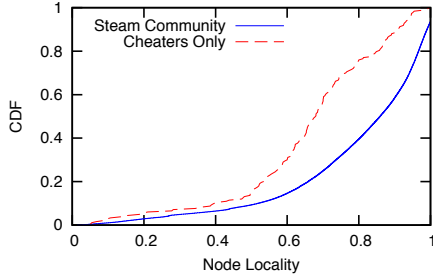


Figure 6: CDF of node locality.

We plot the CDF of node locality for the Steam Community as a whole and just cheaters in Figure 6. We first note that about 40% of Steam Community users have a node locality above 0.90, a phenomenon exhibited by location based social networks BrightKite and FourSquare [4]. Next, when considering only the cheaters embedded within the whole network, we see drastically lower node locality, with only about 10% of cheaters having a node locality greater than 0.90.

From these results, we make two observations: 1) friendships tend to form with other users within close geographic distance, 2) cheaters, while seemingly concentrated in specific geographic regions, tend to have friendships that occur at greater distance than the network as a whole. This might indicate that cheaters form relationships via a different mechanism than non-cheaters.

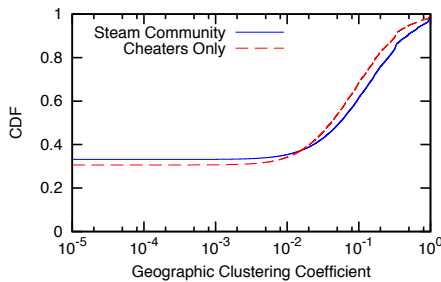


Figure 7: CDF of geographic clustering coefficient.

The CDF for geographic clustering coefficient is plotted in Figure 7. Cheaters embedded within Steam Community tend to have lower geographic clustering coefficients than the network as a whole. Around 10% of users have a geographic clustering coefficient larger than 0.5 with 4% having over 0.9. For cheaters, we see only 5% with a geographic clustering coefficient of over 0.5 and 2% greater than 0.9. While a larger proportion of cheaters have a geographic clustering coefficient over 0.015 than non-cheaters, this trend quickly reverses, with about 40% of the whole network having a

geographic clustering coefficient greater than 0.1 versus 30% of embedded cheaters.

These results might indicate two things. First, cheaters tend to form more geographically dispersed triples when compared to non-cheaters. This follows from the node locality results: if cheaters tend to be further away from their friends than non-cheaters, then triangles that cheaters are involved in might be expected to occur across greater distances than non-cheaters. Second, while we do see evidence of relatively tight geographic clustering, likely due to latency related quality of service concerns inherent to online gaming, we suspect that relationships are formed around community-run servers, thus the distance between triples of users is less important than the distance of triples consisting of two users and a server. This is a hypothesis we are planning to test in future work.

4. SUMMARY

This paper presented an analysis of Steam Community, a large social network of gamers. We built a crawler and discovered over 12 million Steam users connected in an explicitly declared social network. Some of the users have associated permanent cheating flags, and we focus our discussion on these users. Cheaters seem to form the same number of friendships as the network as a whole, and while cheaters tend to own fewer games and play fewer hours than non-cheaters, these metrics increase more acutely for cheaters as the number of friends increases. We find that users are not evenly distributed with respect to location, and that cheaters are even further over-represented in certain regions. Finally, we note that while gamers tend to form geographically close friendships, cheaters' relationships exhibit different geo-social properties, often occurring at greater distances than the rest of the network.

5. ACKNOWLEDGEMENTS

This research was supported by NSF under Grants No. CNS 0952420 and CNS 0831785.

6. REFERENCES

- [1] C. Chambers, W. Feng, S. Sahu, D. Saha, and D. Brandt. Characterizing online games. *IEEE/ACM TON*, 2010.
- [2] 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.
- [3] 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.
- [4] S. Scellato, C. Mascolo, M. Musolesi, and V. Latora. Distance matters: geo-social metrics for online social networks. In *WOSN 10*, 2010.
- [5] M. Szell and S. Thurner. Measuring social dynamics in a massive multiplayer online game. *Social Networks*, 32:313–329, 2010.
- [6] 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.